



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 119787

TO: Michael Borin
Location: rem 2a55
Art Unit: 1631
Thursday, April 22, 2004

2170

Case Serial Number: 10/078090

From: Toby Port
Location: Biotech-Chem Library
Remsen 1A59
Phone: 571-272-2523

toby.port@uspto.gov

Search Notes

Dear Examiner Borin,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port


```

; APPLICANT: SEKI, NAHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 782
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-782

Query Match      25.1%; Score 420.2; DB 16; Length 1718;
Best Local Similarity 95.0%; Pred. No. 4.1e-111;
Matches 434; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY      4 TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 63
Db      1076 TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 1135

QY      64 GCGGTGGGCATCTACGTGTCTCTGTGGTGAACCAAGCTCTAGCTGAGAGGAAGGACTG 123
Db      1136 GCGGTGGGCATCTACGTGTCTCTGTGGTGAACCAAGCTCTAGCTGAGAGGAAGGACTG 1195

QY      124 CGGGTCGGGACCAAGATTCTGCGCGTCAACGACCAATCCCTGGCCCGGTGACCCACGCG 183
Db      1196 CGGGTCGGGACCAAGATTCTGCGCGTCAACGACCAATCCCTGGCCCGGTGACCCACGCG 1255

QY      184 GAGCGCGTCAAGGCTCTGAAGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGG 243
Db      1256 GAGCGCGTCAAGGCTCTGAAGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGG 1315

QY      244 CGCATCCCTGGGGCTACGTACCAACCAATCTACACCTGGTGGTGGACCCCGAGGGCCGC 303
Db      1316 CGCATCTCTGGGGCTACGTACCAACCAATCTACACCTGGTGGTGGACCCCGAGGGCCGC 1375

QY      304 AGCATCTCCCCACCTCGGGCTGCCCCAGCCCCAGCTGGTGGTGGTGGTGGTGGTGGTGG 363
Db      1376 AGCATCTCCCCACCTCGGGCTGCCCCAGCCCCAGCTGGTGGTGGTGGTGGTGGTGGTGG 1435

QY      364 GGTGACCGGAGGAGCACCTCGACCTCTCTGCAAGAGGGGATGAGAAAAGGTGAACCTG 423
Db      1436 GGTGACCGGAGGAGCACCTCGACCTCTCTGCAAGAGGGGATGAGAAAAGGTGAACCTG 1495

QY      424 GTGCTGGGGGACGGCCGCTCCCTGGGCTTCACGATCC 460
Db      1496 GTGGGAAAGGAGGCCAGCCTCTCAGACACCGTATTC 1532

RESULT 5
US-10-029-386-3728
; Sequence 3728, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEONICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```

```

; SEQ ID NO 3728
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
; OTHER INFORMATION: NT HIT: gi14735271, EVALUE 1.00e-123
; OTHER INFORMATION: EST HUMAN HIT: BF969269.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P54817, EVALUE 3.70e-02
US-10-029-386-3728

Query Match      14.4%; Score 240.8; DB 15; Length 509;
Best Local Similarity 77.1%; Pred. No. 3.2e-59;
Matches 293; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY      81 GTCTCTGGTGAACCAAGCTCTCTAGCTGAGAGGAAGGACTGCGGGTCEGGGACCAGAT 140
Db      72 GCCTCTTCTTCCATGTGGGGCTCTTGAGAAAACCAAGGTCTGGGCCAGGGAGGGTGT 131

QY      141 TCTGCGGCTCAACGACAAATCCCTGGCCCGGGTGACCCACGCGGAGGCCGTCAGGCTCT 200
Db      132 CTTCCACACCCCGGACCTCCCTCTCTTAACCTTTGTATCCCTCTCTCCCTAGGCTCT 191

QY      201 GAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGGCCGCGAGCATCTCCCGGGCTA 260
Db      192 GAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGGCCGCGAGCATCTCCCGGGCTA 251

QY      261 CGTACCAACCAACATCTACACCTGGGTGGACCCGCGAGGGCCGCGAGCATCTCCCGGGCTC 320
Db      252 CGTACCAACCAACATCTACACCTGGGTGGACCCGCGAGGGCCGCGAGCATCTCCCGGGCTC 311

QY      321 GGGCTGCCCCAGCCCCACGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 380
Db      312 GGGCTGCCCCAGCCCCACGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 371

QY      381 CTGCACCTCTCTGCAAGAGGGGATGAGAAAAGGTGAACCTGGTGTCTGGGGGACGGCCG 440
Db      372 CTGCACCTCTCTGCAAGAGGGGATGAGAAAAGGTGAAGTGGGTGGGGAAGGAGGCCA 431

QY      441 GTCCTGGGCTTCACGATCC 460
Db      432 GCCTCTCAGACACCGTATTC 451

RESULT 6
US-10-029-386-17428
; Sequence 17428, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEONICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 17428
; LENGTH: 195
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
; OTHER INFORMATION: EST HUMAN HIT: BE890168.1, EVALUE 1.00e-105
; OTHER INFORMATION: NT HIT: gi14735271, EVALUE 1.00e-106
; OTHER INFORMATION: SWISSPROT HIT: P35428, EVALUE 3.70e-01
US-10-029-386-17428

Query Match      11.6%; Score 195; DB 15; Length 195;

```


325	CTGCCCCAGCC	CCCGTGGTGC	CCCTGAGGC	ACAGGAGGTG	ACCGGAGG	AGCACCCCTG	384	
648	GGTTCAACAC	CCCTCC-----	GACACCAGCT	CAGAAAGAT	GGTGTCCGGCGC	ATCGTCCAC	701	
385	CACCTCCTGC	CAAGGAGGGAT	GAGAAAAAG	TGAACCTGGT	GCTGGGGAC	GCGCGGTCC	444	
702	CTATAC-----	-----	-----	-----	-----	-----	728	
445	CTGGGCCTC	ACGATCCGT	GGGGAGCT	GAGTACGG	CCCTTGGC	ATTACATC	ACTGGCGTG	504
729	CTGGGCTTCA	ACATCCGT	GGGGCAAG	GAGTTTGG	CCCTGGGC	ATCTATGT	GTCCTCAAAGTG	788
505	GACCCAGGCT	CTGAAGC	AGGAGG	CAGCGGGCT	CAAGTTGG	GGGACCAG	ATTCTAGAAAGTG	564
789	GACCATGGT	GGGCTGGC	CGAGGAG	AATGGCAT	CAAGGTGG	GGGACCA	GGTCTGGCAGCC	848
565	AATGGCGG	GAGCTTCT	CAACATCT	TACACG	ACGAGGCT	GTCAGGCT	GCTTAAAGTCATCT	624
849	AACGGTGT	CAGGTTTG	ACACATC	AGCCAC	AGCCAG	CGCCGTGG	AGGTGCTGAAGGGCCAA	908
625	CGGCACCT	CATCCTG	ACAGTGA	AGGACG	TCCGGAG	GCTGCCCAT	GCCCGCACCTGTG	684
909	ACGCACAT	CATGCTG	ACCATCA	AGGAGAC	CGGCCGGT	ATCTCTGC	TACAAGGAGATGGTT	968
685	GACGAGAC	CAAGTGG	ATCGCAG	TTCCTCG	GATCAG	GGGAGAC	CAATGGCGAACTCGGCAGGG	744
969	TCTGAGT	ACTGCTG	GCTGGCT	GGACCG	ACTGAG	CAACGG	GGGTGCTGCAGCAGCTGTCCCCGGCC	1028
745	TCTG	748						
1029	TCTG	1032						

```

RESULT 8
US-10-027-632-285026
; Sequence 285026, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 285026
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-285026

```

	Query Match	4.6%;	Score 76.6;	DB 13;	Length 553;
	Best Local Similarity	64.2%;	Pred. No. 2.1e-11;		
	Matches 115;	Conservative	0;	Mismatches 64;	Indels 0; Gaps 0
Qy	1299	TCTGTCGTCTGTCGTCTCTCTCTGTCGTCTGTCGTCTGTCGTCTGTCGTG			1355
Dh	343	TCTCTCTCTCTCTCTCTCTCTCTCTGTCGTCTGTCGTCTGTCGTCTGTCGTG			402

RESULT 15

[illegible]

Search completed: April 22, 2004, 01:59:54
Job time : 707.997 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 19:28:36 ; Search time 2576.09 Seconds
(without alignments)
11222.365 Million cell updates/sec

Title: US-10-078-090-47
Perfect score: 667
Sequence: 1 gcgtggtgcggcccgaggt.....aagcgagaccaggggggaga 667

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	346.2	51.9	132292	9	AL138895	AL138895 Human DNA
C 2	339.2	50.9	171627	2	AF336382	AF336382 Homo sapi
C 3	89.6	13.4	2093	5	BC051776	BC051776 Danio rer
C 4	87.8	13.2	139973	2	AC135372	AC135372 Rattus no
C 5	87.8	13.2	214625	10	AC116582	AC116582 Mus muscu
C 6	86.6	13.0	211588	2	BX537348	BX537348 Danio rer
C 7	86.2	12.9	46643	10	AL732512	AL732512 Mouse DNA
C 8	85.2	12.8	166893	9	AP002987	AP002987 Homo sapi
C 9	85.2	12.8	179355	10	AC115631	AC115631 Mus muscu
C 10	85.2	12.8	213715	2	BX649528	BX649528 Danio rer
C 11	85	12.7	203271	2	BX322566	BX322566 Danio rer
C 12	85	12.7	215778	2	BX530074	BX530074 Danio rer
C 13	84.8	12.7	135904	2	AC139157	AC139157 Mus muscu
C 14	84.8	12.7	213673	10	AC130218	AC130218 Mus muscu
C 15	84.6	12.7	117886	2	AC139385	AC139385 Rattus no
C 16	84.6	12.7	160718	2	BX664714	BX664714 Danio rer
C 17	84.6	12.7	202878	2	BX511101	BX511101 Danio rer
C 18	84.6	12.7	228962	2	AC131587	AC131587 Mus muscu
C 19	84.6	12.7	229574	2	AC123856	AC123856 Mus muscu
C 20	84.4	12.7	181884	2	AC121089	AC121089 Mus muscu
C 21	84.4	12.7	205170	2	AC113482	AC113482 Mus muscu
C 22	84.2	12.6	137393	2	BX545852	BX545852 Danio rer
C 23	84	12.6	169095	5	BX088685	BX088685 Zebrafish
C 24	84	12.6	230877	5	BX248500	BX248500 Zebrafish
C 25	84	12.6	237343	2	BX569792	BX569792 Danio rer
C 26	83.8	12.6	165993	5	AL772362	AL772362 Zebrafish
C 27	83.6	12.5	142803	2	AC013771	AC013771 Homo sapi
C 28	83.6	12.5	169086	2	AC015832	AC015832 Homo sapi
C 29	83.6	12.5	169821	2	AC019098	AC019098 Homo sapi
C 30	83.6	12.5	199495	9	AP003441	AP003441 Homo sapi
C 31	83.6	12.5	247113	2	AC107008	AC107008 Rattus no
C 32	83.6	12.5	256519	2	AC115517	AC115517 Rattus no
C 33	83.4	12.5	130176	2	BX005295	BX005295 Danio rer
C 34	83.4	12.5	171807	2	BX000463	BX000463 Danio rer
C 35	83.4	12.5	183326	2	BX571886	BX571886 Danio rer
C 36	83.2	12.5	263668	2	AC136184	AC136184 Rattus no
C 37	83.2	12.5	323010	2	AC111643	AC111643 Rattus no
C 38	83	12.4	127560	5	BX005077	BX005077 Zebrafish
C 39	83	12.4	179186	10	AC125396	AC125396 Mus muscu
C 40	83	12.4	196887	2	AC114586	AC114586 Mus muscu
C 41	83	12.4	200207	2	BX323062	BX323062 Danio rer
C 42	83	12.4	201642	10	AL672174	AL672174 Mouse DNA
C 43	82.8	12.4	216853	2	BX284112	BX284112 Danio rer
C 44	82.6	12.4	110000	2	BX572080_0	BX572080 Danio rer
C 45	82.6	12.4	110000	2	BX572080_1	Continuation (2 of

ALIGNMENTS

RESULT 1
AL138895/c AL138895 132292 bp DNA linear PRI 27-JUN-2001
LOCUS Human DNA sequence from clone Rp11-9M16 on chromosome 9, complete
DEFINITION sequence.
ACCESSION AL138895
VERSION AL138895.16 GI:14575083
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 132292)
AUTHORS Skuce, C.
TITLE Direct Submission

JOURNAL Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Jun 28, 2001 this sequence version replaced gi:14456172. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-9M16 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-9M16 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-9M16 is at 132292 in this sequence. The true left end of clone RP11-402G3 is at 124695 in this sequence. The true right end of clone RP11-8211 is at 2000 in this sequence.

FEATURES

Source Location/Qualifiers

1..132292

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="9"

/clone="RP11-9M16"

/clone_lib="RPCI-11.1"

1..32

/note="AluY repeat: matches 1..32 of consensus"

74..301

/note="L1MC1 repeat: matches 5718..5955 of consensus"

399..696

/note="AluSq repeat: matches 1..298 of consensus"

1186..1273

/note="MIR repeat: matches 49..142 of consensus"

2271..2435

/note="MER63A repeat: matches 46..206 of consensus"

2467..2553

/note="MIR repeat: matches 79..165 of consensus"

2587..2713

/note="L2 repeat: matches 2622..2740 of consensus"

2911..3207

/note="AluSc repeat: matches 5..301 of consensus"

3214..3262

/note="MIR repeat: matches 115..166 of consensus"

4193..4262

/note="MIR repeat: matches 60..130 of consensus"

4368..5308

/note="CpG island"

/evidence=not_experimental

5299..5606

/note="AluSx repeat: matches 12..312 of consensus"

5686..5787

/note="MIR repeat: matches 141..262 of consensus"

6714..6799

repeat_region /note="MIR repeat: matches 20..114 of consensus" 6995..7437

repeat_region /note="MLT1D repeat: matches 3..503 of consensus" 7440..7502

repeat_region /note="MIR repeat: matches 82..144 of consensus" 8034..8179

repeat_region /note="LTR16A repeat: matches 200..354 of consensus" 8160..8271

repeat_region /note="LTR16C repeat: matches 255..368 of consensus" 10037..10131

repeat_region /note="MIR repeat: matches 14..123 of consensus" 10604..10711

repeat_region /note="L2 repeat: matches 2405..2521 of consensus" 11151..11327

repeat_region /note="MIR repeat: matches 18..205 of consensus" 11579..11883

repeat_region /note="AluSx repeat: matches 1..304 of consensus" 11886..12024

repeat_region /note="MIR repeat: matches 97..243 of consensus" 12598..12717

repeat_region /note="MIR repeat: matches 52..181 of consensus" 13839..14002

repeat_region /note="MIR repeat: matches 80..258 of consensus" 14071..14205

repeat_region /note="MIR repeat: matches 92..226 of consensus" 15216..15653

repeat_region /note="MLT2CB repeat: matches 1..492 of consensus" 15702..15995

repeat_region /note="L2 repeat: matches 2372..2688 of consensus" 16113..16847

repeat_region /note="L2 repeat: matches 1940..2707 of consensus" 16875..16922

repeat_region /note="L2 copies 4 mer atga 81% conserved" 17482..17573

repeat_region /note="MIR repeat: matches 125..230 of consensus" 18163..18379

repeat_region /note="MIR repeat: matches 23..241 of consensus" 18972..19080

repeat_region /note="MIR repeat: matches 34..144 of consensus" 20542..20994

repeat_region /note="Charlie4a repeat: matches 26..506 of consensus" 21421..21521

repeat_region /note="L2 repeat: matches 2045..2158 of consensus" 21523..21758

repeat_region /note="AluJo repeat: matches 1..232 of consensus" 22863..23243

repeat_region /note="MSTA repeat: matches 1..426 of consensus" 23319..23495

repeat_region /note="MIR repeat: matches 5..192 of consensus" 24438..24542

repeat_region /note="MIR repeat: matches 28..140 of consensus" 24993..25526

repeat_region /note="L2 repeat: matches 2099..2709 of consensus" 25630..25867

repeat_region /note="MIR repeat: matches 16..262 of consensus" 25889..26233

repeat_region /note="L2 repeat: matches 2363..2674 of consensus" 26257..26381

repeat_region /note="MIR repeat: matches 28..153 of consensus" 26919..27000

repeat_region /note="MIR repeat: matches 64..147 of consensus" 27292..27435

repeat_region /note="MIR repeat: matches 102..254 of consensus" 29060..29358

repeat_region /note="AluSx repeat: matches 1..298 of consensus" 29523..29664

repeat_region /note="L2 repeat: matches 2593..2734 of consensus" 29669..29777

repeat_region /note="MIR repeat: matches 17..140 of consensus" 30134..30243

repeat_region /note="L2 repeat: matches 2167..2289 of consensus" 30460..30627

repeat_region /note="MER5B repeat: matches 1..173 of consensus"


```

1. .2093
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="MGC:56117 IMAGE:5411459"
/tissue_type="Whole body, adult male"
/clone_lib="Sugano SJD adult male"
/lab_host="DH10B"
/note="Vector: pME18S-FL3"
43. .555
/codon_start=1
/product="Unknown (protein for MGC:56117)"
/protein_id="AAH51776.1"
/db_xref="GI:30353804"
/translation="MAETKADLTADASNNNNVSSQNNPLSRKLHKILETRLNDNDK
EMLEALKSLSVFFTSNLRTRNLGRDIERRSLSINEEFVRIFKDVKEELESVHEDVQ
AMSSCCSEMTNRLKAAKEQTQDLIVKTNKLQGENHRLVRAQVAQFLSKFQLSAAEM
ATLRSAPQDP"

```

```

/codon_start=1
/product="Unknown (protein for MGC:56:17)"
/protein_id="AAH5176.1"
/db_xref="GI:30353804"
/translation="MAETKADLTADASNSSNNNVSSQNNPLSRKLHKILETRLDNDK
EMLEAKLSVFFTESSLRTNRNLRGDIERRSLSINEEFVRIFKDVKEELESVHEDVQ
AMSSCCEEMTRLKAAKEQTQDLIVKTNKLGENHRLVRAQVAQFLSKFQLSAAEM
ATLRSAPQDP"

```

[illegible]

RESULT 4
AC135372
LOCUS
AC135372 linear HTG 13-OCT-2002
DEFINITION
Rattus norvegicus clone CH230-329K10, *** SEQUENCING IN PROGRESS
***, 66 unordered pieces.

LOCUS AC135372 139973 bp DNA linear HTG 13-OCT-2002
DEFINITION Rattus norvegicus clone CH230-329K10, *** SEQUENCING IN PROGRESS
***, 66 unordered pieces.
ACCESSION AC135372
VERSION AC135372.1 GI:23928474
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 139973)
AUTHORS Muzny, D., Marie, J., Metzker, M., Lee, J., Adams, C., Alder, J.,

REFERENCE
AUTHORS

Rattus.
1 (bases 1 to 139973)
Muzny, D., Marie, Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Mirja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G.,
Olarunpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabore,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wleczyk,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 139973)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: QOSY
Center clone name: CH230-329K10
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 65209 bases at least Q40
Consensus quality: 71948 bases at least Q30
Consensus quality: 77254 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 66 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1002: contig of 1002 bp in length
1003 1102: gap of unknown length
1103 2267: contig of 1165 bp in length
2268 2367: gap of unknown length
2368 3779: contig of 1412 bp in length
3780 3879: gap of unknown length
3880 5068: contig of 1189 bp in length
5069 5168: gap of unknown length
5169 6321: contig of 1153 bp in length

6322 *
6422 *
7507 *
7607 *
9198 *
9298 *
10744 *
10844 *
11872 *
11972 *
13022 *
13122 *
14341 *
14441 *
15455 *
15555 *
16846 *
16946 *
18081 *
18181 *
19692 *
19792 *
19793 *
21511 *
21611 *
22628 *
22728 *
23849 *
23949 *
25019 *
25119 *
26231 *
26331 *
27734 *
27834 *
29270 *
29370 *
30593 *
30693 *
31886 *
31986 *
33220 *
33320 *
35165 *
35265 *
36760 *
36860 *
38181 *
38281 *
39504 *
39604 *
41381 *
41481 *
42887 *
42987 *
44736 *
44836 *
46343 *
46443 *
48465 *
48565 *
49990 *
50090 *
51488 *
51588 *
53758 *
53858 *
55206 *
55306 *
57183 *
57283 *
58631 *
58731 *
60122 *
60221: gap of unknown length
6421: gap of unknown length
7506: contig of 1085 bp in length
7606: gap of unknown length
9197: contig of 1591 bp in length
9297: gap of unknown length
10743: contig of 1446 bp in length
10843: gap of unknown length
11871: contig of 1028 bp in length
11971: gap of unknown length
13021: contig of 1050 bp in length
13121: gap of unknown length
14340: contig of 1219 bp in length
14440: gap of unknown length
15454: contig of 1014 bp in length
15554: gap of unknown length
16845: contig of 1291 bp in length
16945: gap of unknown length
18080: contig of 1135 bp in length
18180: gap of unknown length
19692: contig of 1512 bp in length
19792: gap of unknown length
21510: contig of 1718 bp in length
21610: gap of unknown length
22627: contig of 1017 bp in length
22727: gap of unknown length
23848: contig of 1121 bp in length
23948: gap of unknown length
25018: contig of 1070 bp in length
25118: gap of unknown length
26230: contig of 1112 bp in length
26330: gap of unknown length
27733: contig of 1403 bp in length
27833: gap of unknown length
29269: contig of 1436 bp in length
29369: gap of unknown length
30592: contig of 1223 bp in length
30692: gap of unknown length
31885: contig of 1193 bp in length
31985: gap of unknown length
33219: contig of 1234 bp in length
33319: gap of unknown length
35164: contig of 1845 bp in length
35264: gap of unknown length
36759: contig of 1495 bp in length
36859: gap of unknown length
38180: contig of 1321 bp in length
39503: contig of 1223 bp in length
39603: gap of unknown length
41380: contig of 1777 bp in length
41480: gap of unknown length
42886: contig of 1406 bp in length
42986: gap of unknown length
44735: contig of 1749 bp in length
44835: gap of unknown length
46342: contig of 1507 bp in length
46442: gap of unknown length
48464: contig of 2022 bp in length
48564: gap of unknown length
49989: contig of 1425 bp in length
50089: gap of unknown length
51487: contig of 1398 bp in length
51587: gap of unknown length
53757: contig of 2170 bp in length
53857: gap of unknown length
55205: contig of 1348 bp in length
55305: gap of unknown length
57182: contig of 1877 bp in length
57282: gap of unknown length
58630: contig of 1348 bp in length
58730: gap of unknown length
60121: contig of 1391 bp in length
60221: gap of unknown length

[illegible]

[illegible]

```

RESULT 12
BX530074/c
LOCUS
DEFINITION
BX530074
BX530074
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
BX530074
BX530074
215778 bp
DNA
linear
HTG 29-SEP-2003
Danio rerio clone DKEY-266K15, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
BX530074
BX530074
GI:37051046
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1
(bases 1 to 215778)
Mclaren,S.
Direct Submission
Submitted (28-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 29, 2003 this sequence version replaced gi:32134827.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
----- Project Information
Center project name: zK266K15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 213339 bases at least Q40
Consensus quality: 214215 bases at least Q30
Consensus quality: 214727 bases at least Q20
Insert size: 215178; sum-of-contigs
Insert size: 198791; 5.1% error; agarose-fp
Quality coverage: 7.70x in Q20 bases; sum-of-contigs Quality
coverage: 8.44x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4918: contig of 4918 bp in length
* 4919 5018: gap of 100 bp
* 5019 71913: contig of 66895 bp in length
* 71914 72013: gap of 100 bp
* 72014 125787: contig of 53774 bp in length
* 125788 125887: gap of 100 bp
* 125888 189725: contig of 63838 bp in length
* 189726 189825: gap of 100 bp
* 189826 191911: contig of 2086 bp in length
* 191912 192011: gap of 100 bp
* 192012 206777: contig of 14766 bp in length
* 206778 206877: gap of 100 bp
* 206878 215778: contig of 8901 bp in length.
*
* Location/Qualifiers
* 1. .215778
* /organism="Danio rerio"
* /mol_type="genomic DNA"
* /db_xref="taxon:7955"
* /clone="DKEY-266K15"
* /clone_lib="DanioKey"
*
* misc_feature
* 1. .4918
* /note="assembly fragment:01270
* fragment chain:1"
* 5019. .71913
* /note="assembly fragment:02287
*
FEATURES
source
misc_feature
misc_feature

```


Db 83170 ACAGCTTTAAATTCAAATTTTCAAAGATATCTATAACATCAGAAACTAAA 83119

RESULT 14

AC130218 213673 bp DNA linear ROD 19-NOV-2003

LOCUS Mus musculus chromosome 18 clone RP23-121N18, complete sequence.

DEFINITION AC130218

ACCESSION AC130218.4 GI:38424199

VERSION HTG.

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Mus musculus

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE 1 (bases 1 to 213673)

JOURNAL Wilson,R.K.

REFERENCE

AUTHORS The sequence of Mus musculus clone

TITLE Unpublished

JOURNAL

REFERENCE

AUTHORS 2 (bases 1 to 213673)

TITLE McPherson,J.D. and Waterston,R.H.

JOURNAL Direct Submission

REFERENCE

AUTHORS Submitted (08-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

TITLE 3 (bases 1 to 213673)

JOURNAL McPherson,J.D. and Waterston,R.H.

REFERENCE

AUTHORS Direct Submission

TITLE Submitted (28-FEB-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

JOURNAL 4 (bases 1 to 213673)

REFERENCE

AUTHORS Wilson,R.K.

TITLE Direct Submission

JOURNAL Submitted (19-NOV-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Nov 19, 2003 this sequence version replaced gi:28604207.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu>

Contact: submissions@watson.wustl.edu

----- Project Information -----

Center project name: M_BA0121N18

----- Location/Qualifiers -----

1. 213673

/organism="Mus musculus"

/mol type="genomic DNA"

/db xref="taxon:10090"

/chromosome="18"

/clone="RP23-121N18"

ORIGIN

Query Match 12.7%; Score 84.8; DB 10; Length 213673;

Best Local Similarity 60.3%; Pred. No. 1.3e-11;

Matches 140; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 236 CAGAAATCTCAGCCTGGTGGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCT 295

Db 56329 CAGAACAGGATCCCTGTACTGGATCTTGAAGAACTCATTGTGTGTGTGTGTGTGTGT 56388

QY 296 GTCTGTCTGTCTCTGT 355

Db 56389 GT 56448

QY 356 GTGTGTGTCTCCTCACTCTTTCATCCCTATCATAGTAGTATGATATAATAAATATTAG 415

Db 56449 GTGTGTGTATATATATATATTTATCTATAGCCGTAAAGTAAAAAGGTTAAAAAATGC 56508

QY 416 AGAGATACACAGAAAAATATATAGAGAGATAACAGTGTCTCTATATAAAAAA 467

Db 56509 ACAGCTTTAAATTCAAATTTTCAAAGATATCTATAACATCAGAAACTAAA 56560

RESULT 15

AC139385/c

LOCUS

DEFINITION

AC139385

ACCESSION

VERSION AC139385.5 GI:30521076

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

TITLE 1 (bases 1 to 117886)

JOURNAL Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Wilson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 117886)

Worley,K.C.

Direct Submission

Submitted (02-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 117886)

Worley,K.C.

Direct Submission

AC139385 117886 bp DNA linear HTG 10-MAY-2003

Rattus norvegicus clone CH230-63O14, WORKING DRAFT SEQUENCE, 26 unordered pieces.

AC139385

AC139385

AC139385.5 GI:30521076

HTG; HTGS_PHASE1; HTGS_DRAFT.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 117886)

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Wilson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 117886)

Worley,K.C.

Direct Submission

Submitted (02-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 117886)

Worley,K.C.

Direct Submission

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 19:26:07 ; Search time 651.769 Seconds
(without alignments)
10930.600 Million cell updates/sec

Title: US-10-078-090-48

Perfect score: 1677

Sequence: 1 gagtgcgcgcgtgccaaagc.....aagcgagaccaggggggaga 1677

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1677	100.0	1677	6	ABV83605	Abv83605 Human bre
2	647.2	38.6	667	6	ABV83604	Abv83604 Human bre
3	554.8	33.1	3102	3	AAC75441	Aac75441 Human ORF
4	420.2	25.1	1718	7	ADA53214	Ada53214 Human cod
5	108.6	6.5	1338	4	ABL29731	Abi29731 Drosophil
6	108.6	6.5	6497	4	ABL29730	Abi29730 Drosophil
7	107.6	6.4	2822	7	ABX71178	Abx71178 Novel hum
8	76.8	4.6	3316	4	ABL10742	Abi10742 Drosophil
9	76.6	4.6	1817	5	ABA15256	Abai5256 Human ner
10	75.6	4.5	271990	9	ADD25213	Add25213 Fertility
11	75.2	4.5	710	6	ABT09598	Abt09598 Phase-1 R
12	73.8	4.4	1274	4	AAH73917	Aah73917 Human PDZ
13	73.6	4.4	2278	4	ABL12518	Abi12518 Drosophil
14	73.4	4.4	667	9	ADD19798	Add19798 Seabass p
15	73.4	4.4	67832	8	ADA02801	Ada02801 Mouse Itk
16	73.4	4.4	67832	9	ADB72539	Adb72539 Mouse Itk
17	73.4	4.4	67832	9	ADC85281	Adc85281 Mouse Itk
18	73.2	4.4	42998	7	ADA14747	Ada14747 Human rib
19	73.2	4.4	42999	6	ABS65032	Abs65032 Invertebr
20	73.2	4.4	42999	9	AAD61411	Aad61411 Human rib
21	73	4.4	742	6	ABS76748	Abs76748 Frog embr
22	72.6	4.3	338	9	ADD19844	Add19844 Oreochrom
23	72.4	4.3	596	9	ADD19952	Add19952 Oreochrom

C	24	72.4	4.3	814	9	ADD20177	Add20177 Oreochrom
	25	72.4	4.3	41907	8	ADA37416	Ada37416 Human rDN
C	26	72.2	4.3	534	9	ADD20617	Add20617 Oreochrom
C	27	72.2	4.3	110000	7	ABX16390_5	Continuation (6 of
C	28	72	4.3	448	2	AAT68682	Aat68682 Penaeus v
C	29	72	4.3	605	9	ADD20225	Add20225 Oreochrom
	30	72	4.3	96599	9	ADC85298	Adc85298 Human Egr
	31	72	4.3	96600	8	ADA02819	Ada02819 Mouse Sos
	32	72	4.3	96600	9	ADB72557	Adb72557 Mouse Sos
C	33	71.8	4.3	299	9	ADD19778	Add19778 Gadus mor
	34	71.8	4.3	475	4	AAI12688	Aai12688 Probe #26
	35	71.8	4.3	475	4	ABA54378	Aba54378 Human foe
	36	71.8	4.3	475	4	AAI34039	Aai34039 Probe #27
	37	71.8	4.3	475	4	ABA43925	Aba43925 Human bre
	38	71.8	4.3	475	4	ABA24157	Aba24157 Probe #26
	39	71.8	4.3	475	4	AAK28111	Aak28111 Human bon
	40	71.8	4.3	475	4	AAK02673	Aak02673 Human bra
	41	71.8	4.3	475	4	ABS27717	Abs27717 Human liv
	42	71.8	4.3	475	5	AAI02597	Aai02597 Probe #25
	43	71.8	4.3	475	6	ABS02604	Abs02604 Human gen
C	44	71.8	4.3	527	6	ABQ92801	Abq92801 Triticum
	45	71.8	4.3	546	4	AAI19220	Aai19220 Probe #91

ALIGNMENTS

RESULT 1

ABV83605

ID ABV83605 standard; cDNA; 1677 BP.

XX AC ABV83605;

XX AC ABV83605;

DT 06-DEC-2002 (first entry)

XX Human breast specific gene SEQ ID NO 48.

DE Human; breast; cytostatic; cancer; transgenic; gene therapy; vaccine;

XX gene; ss.

XX Homo sapiens.

XX WO200266605-A2.

XX 29-AUG-2002.

PF 14-FEB-2002; 2002WO-US004284.

XX 15-FEB-2001; 2001US-0268999P.

XX (DIAD-) DIADEXUS INC.

XX Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;

PI Sun Y, Liu C;

XX WPI; 2002-713345/77.

XX New isolated breast specific nucleic acid molecules and polypeptides, useful for identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue.

PS Claim 1; Page 171-172; 254pp; English.

XX The invention relates to human breast specific nucleic acids (I)

CC comprising: (a) a sequence encoding any one of 95 protein sequences (ABP66614-ABP66708); (b) any one of 115 polynucleotide sequences (ABV83558-ABV83672); (c) a molecule that selectively hybridizes to (a) or (b); (d) a molecule having at least 60% sequence identity to (a) or (b). The breast specific nucleic acid molecules, polypeptides and antibodies are useful for identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue. They are also useful for producing transgenic animals and cells and producing engineered breast tissue for treatment and research. The

CC	disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC	diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC	storage, systemic lupus erythematosus, severe combined immunodeficiency
CC	(SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC	disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC	cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC	enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX	
5Q	Sequence 3102 BP; 699 A; 996 C; 860 G; 545 T; 0 U; 2 Other;
	Query Match 33.1%; Score 554.8; DB 3; Length 3102;
	Best Local Similarity 99.6%; Pred. No. 9e-130;
	Matches 556; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	190 GTCAAGGCTCTGAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGGCGCATC 249
DB	
QY	250 CCTGGGGGCTACGTACCAACCATCTACACCTGGGTGGACCCGACGGGCCGCAGCATC 309
DB	
QY	310 TCCCCACCTCGGGCCTGCCACGCCCCACGGTGGTGCCTGAGGCACGAGGGGTGAC 369
DB	
QY	370 CGGAGGAGCACCTTGCACTCTCGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTCTG 429
DB	
QY	430 GGGGACGGCCGTCCCTGGGCCCTCACGATCCGTGGGGAGCTGATACGGCCTTGGCATT 489
DB	
QY	490 TACATCACTGGCGTGGAACCCAGGCTCTGAAGCAGAGGCAGCGGGCTCAAGGTTGGGAC 549
DB	
QY	550 CAGATTCTAGAAGTGAATGGGGGGAGCTTTCTCAACATCCTACACGACGAGGCTGTCAAG 609
DB	
QY	610 CTGCTTAAGTCATCTCGGCACCTCATCTGTACAGTGAAGGACGTCGGGAGGCTGCCCCAT 669
DB	
QY	670 GCCCGACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCCTCCGGATCAGGGAGCCATG 729
DB	
QY	742 GCGCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCCTCCGGATCAGGGAGCCATG 801
DB	
QY	730 GCGAACTCGGCAGGGTCT 747
DB	
QY	802 GCGAACTCGGCAGGGTTT 819
DB	

19-MAR-2003.
21-MAR-2002; 2002EP-00006586.
14-SEP-2001; 2001JP-00328381.
24-JAN-2002; 2002US-0350435P.
(HELI-) HELIX RES INST.
(REAS-) RES ASSOC BIOTECHNOLOGY.
Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
WPI; 2003-395539/38.
P-PSDB; ADA54853.
New polynucleotides encoding full-length polypeptides, e.g. secretory
and/or membrane proteins, useful for developing medicines for diseases in
which the gene is involved, or as target molecules for gene therapy.
Claim 1; SEQ ID NO 782; 205pp; English.
The present invention relates to novel human secretory or membrane
proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
ADA54071). The coding sequences are useful in the gene therapy of
diseases caused by abnormalities of the proteins, e.g. cancer,
inflammatory diseases, osteoporosis or neurological disease.
Sequence 1718 BP; 269 A; 611 C; 551 G; 287 T; 0 U; 0 Other;

Query Match	25.1%;	Score 420.2;	DB 7;	Length 1718;
Best Local Similarity	95.0%;	Pred. No. 8.1e-96;		
Matches 434;	Conservative	0;	Mismatches 23;	Indels 0;
				Gaps 0;

QY	4	TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC	63
Db	1076	TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC	1135
QY	64	GGCGTGGGCATCTACGTGTCTCTGGTGGAAACGAGGCTCTCTAGCTGAGAAAGGACTG	123
Db	1136	GGCGTGGGCATCTACGTGTCTCTGGTGGAAACGAGGCTCTCTAGCTGAGAAAGGACTG	1195
QY	124	CGGTCGGGGACCAAGATTCTGCGCTCAACGACAAATCCCTGGCCCGGTGACCCACGCG	183
Db	1196	CGGTCGGGGACCAAGATTCTGCGCTCAACGACAAATCCCTGGCCCGGTGACCCACGCG	1255
QY	184	GAGGCGCTCAAGGCTCTGAAGGCTTCAAGAGCTGGTGTCTGTGTACTCAGCAGGG	243
Db	1256	GAGGCGCTCAAGGCTCTGAAGGCTTCAAGAGCTGGTGTCTGTGTACTCAGCAGGG	1315
QY	244	CGCATCCCTGGGGCTACGTCAACCAACACATCTACACCTGGGTGGACCCGACGGCCCGC	303
Db	1316	CGCATCTCTGGGGCTACGTCAACCAACACATCTACACCTGGGTGGACCCGACGGCCCGC	1375
QY	304	AGCATCTCCCAACCTCGGGCCTGCCCGACCCACCGTGGTGCCCTGAGGCAGCAGGAG	363
Db	1376	AGCATCTCCCAACCTCGGGCCTGCCCGACCCACCGTGGTGCCCTGAGGCAGCAGGAG	1435
QY	364	GGTGACCGGAGGAGCACCTTGCACCTCTCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTG	423
Db	1436	GGTGACCGGAGGAGCACCTTGCACCTCTCTGCAAGGAGGGGATGAGAAAAAGGTGAGTGGG	1495
QY	424	GTGCTGGGGACGGCCGGTCCCTGGGGCCTCAGCATCC	460
Db	1496	GTGGGAAAGGAGGCCAGCCTCTCAGACACCGTATTTC	1532

RESULT 5
ABL29731
ID ABL29731 standard; DNA; 1338 BP.
XX
AC ABL29731;

RESULT 4	
ADA53214	
ID	ADA53214 standard; cDNA; 1718 BP.
XX	
AC	ADA53214;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Human coding sequence, SEQ ID 782.
XX	
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW	inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1293569-A2.
XX	

XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 40666.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 40666; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1338 BP; 303 A; 400 C; 384 G; 251 T; 0 U; 0 Other;
Query Match 6.5%; Score 108.6; DB 4; Length 1338;
Best Local Similarity 60.8%; Pred. No. 5.4e-17;
Matches 177; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
QY 415 GTGAACCTGTGCTGGGGACGGCCGCTCCCTGGCCCTCACGATCCGTGGGGAGCTGAG 474
Db 268 GTGGAGCTGCTGATCGAACCTGGTCAGTCTCTGGCCCTGATGATCCGTGGCGGTGGAG 327
QY 475 TACGGCCTTGGCATTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGCGGCGG 534
Db 328 TACGGCCTGGGGATCTTTGTACCCGGCGTAGATAAGGACAGCGTGGCAGATCGATCCGGA 387
QY 535 CTCAAGGTTGGGACCCAGATTCTAGAGTGAATGGGCGGAGCTTTCTCAACATCCTACAC 594
Db 388 CTGATGATTGGCAGCAGAGATCCTCGAGGTCAATGGGCAATCCITTCGATGTGACGCAC 447
QY 595 GACGAGGCTGTACGGCTGCTTAAGTCACTCTCGGCACCTCATCTCTGACAGTGAAGGACGTC 654
Db 448 GACGAGGCGGTGGGTGAGTTGAAGTACCACAAACGCATGTCGCTGGTGATACGTGACGTG 507
QY 655 GGGAGGCTGCCCATGCCCGCACCTGTGGACGAGACCAAGTGGATCGCC 705
Db 508 GGCAAGGTGCCCCACTCTGCACCTCCATCGAGATGGAGCCCTGGGACGCC 558

RESULT 6
ABL29730/c
ID ABL29730 standard; DNA; 6497 BP.
XX

AC ABL29730;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 40663.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 40663; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6497 BP; 1933 A; 1221 C; 1417 G; 1926 T; 0 U; 0 Other;
Query Match 6.5%; Score 108.6; DB 4; Length 6497;
Best Local Similarity 60.8%; Pred. No. 1.1e-16;
Matches 177; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
QY 415 GTGAACCTGTGCTGGGGACGGCCGCTCCCTGGCCCTCACGATCCGTGGGGAGCTGAG 474
Db 4982 GTGGAGCTGCTGATCGAACCTGGTCAGTCTCTGGCCCTGATGATCCGTGGCGGTGGAG 4923
QY 475 TACGGCCTTGGCATTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGCGGCGG 534
Db 4922 TACGGCCTGGGGATCTTTGTACCCGGCGTAGATAAGGACAGCGTGGCAGATCGATCCGGA 4863
QY 535 CTCAAGGTTGGGACCCAGATTCTAGAGTGAATGGGCGGAGCTTTCTCAACATCCTACAC 594
Db 4862 CTGATGATTGGCGACGAGATCCTCGAGGTCAATGGGCAATCCTTTCTCGATGTGACGCAC 4803
QY 595 GACGAGGCTGTACGGCTGCTTAAGTCACTCTCGGCACCTCATCTGACAGTGAAGGACGTC 654
Db 4802 GACGAGGCGGTGGGTGAGTTGAAGTACCACAAACGCATGTCGCTGGTGATACGTGACGTG 4743
QY 655 GGGAGGCTGCCCATGCCCGCACCTGTGGACGAGACCAAGTGGATCGCC 705
Db 4742 GGCAAGGTGCCCCACTCTGCACCTCCATCGAGATGGAGCCCTGGGACGCC 4692

RESULT 7
ABL29730/c
ID ABX71178 standard; cDNA; 2822 BP.
XX

ABX711178;
05-MAR-2003 (first entry)
Novel human cDNA sequence #403.
Human; gene; ss; nervous system disorder; peripheral neuropathy;
Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
Crohn's disease; anaphylaxis; proliferation; chemotactic;
differentiation; stem cell growth factor; haematopoiesis; chemoketic;
haemostatic; antiinflammatory; expressed sequence tag; EST.
Homo sapiens.
WO200281731-A2.
17-OCT-2002.
29-JAN-2002; 2002WO-US001222.
30-JAN-2001; 2001US-00774528.
(HYSE-) HYSEQ INC.
(GOOD/) GOODRICH R W.
Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
WPI; 2003-058563/05.
Novel polypeptide useful for treating neurodegenerative diseases, myeloid
or lymphoid cell disorders, bone disorders, mechanical and traumatic
disorders, coagulation disorders, and inflammatory diseases.
Claim 1; Page; 612pp; English.
This invention relates to the cDNA sequences encoding an isolated novel
human polypeptide. The protein encoded by the nucleic acid of the
invention is useful for treating central and peripheral nervous system
diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
Alzheimer's disease); autoimmune disease (e.g. systemic lupus
erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)
; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
trauma); lung or liver fibrosis; reperfusion injury in various tissues;
bacterial, viral or fungal infections; allergic conditions such as
allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
disease, anaphylaxis). The protein may be used to inhibit the growth,
infection or function of infectious agents such as bacteria, fungi,
viruses, or to effect bodily characteristics, biorhythms or circadian
cycles of rhythms. The protein may also have
proliferation/differentiation, stem cell growth factor, haematopoiesis
regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
activities. The cDNA sequences of the invention are useful for expressing
recombinant protein for analysis. The present sequence represents a novel
human cDNA sequence of the invention, this sequence is an expressed
sequence tag (EST) and was identified using subtractive hybridisation

**END OF
CCT
CCT
CCT
CCT
CCT
CCT**

PR	24-FEB-2000;	2000US-01846564P.
PR	02-MAR-2000;	2000US-01863505P.
PR	16-MAR-2000;	2000US-01898742P.
PR	17-MAR-2000;	2000US-01900767P.
PR	18-APR-2000;	2000US-01981233P.
PR	19-MAY-2000;	2000US-0205515P.
PR	07-JUN-2000;	2000US-0209467P.
PR	28-JUN-2000;	2000US-0214886P.
PR	30-JUN-2000;	2000US-0215135P.
PR	07-JUL-2000;	2000US-0216647P.
PR	07-JUL-2000;	2000US-0216880P.
PR	11-JUL-2000;	2000US-0217487P.
PR	11-JUL-2000;	2000US-0217496P.
PR	14-JUL-2000;	2000US-0218290P.
PR	26-JUL-2000;	2000US-0220963P.
PR	14-AUG-2000;	2000US-0224518P.
PR	14-AUG-2000;	2000US-0224519P.
PR	14-AUG-2000;	2000US-0225213P.
PR	14-AUG-2000;	2000US-0225214P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225447P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0226868P.
PR	22-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227009P.
PR	30-AUG-2000;	2000US-0228924P.
PR	01-SEP-2000;	2000US-0229287P.
PR	01-SEP-2000;	2000US-0229343P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	06-SEP-2000;	2000US-0230438P.
PR	08-SEP-2000;	2000US-0231242P.
PR	08-SEP-2000;	2000US-0231243P.
PR	08-SEP-2000;	2000US-0231244P.
PR	08-SEP-2000;	2000US-0231413P.
PR	08-SEP-2000;	2000US-0231414P.
PR	08-SEP-2000;	2000US-0232080P.
PR	08-SEP-2000;	2000US-0232081P.
PR	12-SEP-2000;	2000US-0231968P.
PR	14-SEP-2000;	2000US-0232397P.
PR	14-SEP-2000;	2000US-0232398P.
PR	14-SEP-2000;	2000US-0232399P.
PR	14-SEP-2000;	2000US-0232400P.
PR	14-SEP-2000;	2000US-0232401P.
PR	14-SEP-2000;	2000US-0233063P.
PR	14-SEP-2000;	2000US-0233064P.
PR	14-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234997P.
PR	25-SEP-2000;	2000US-0234998P.
PR	26-SEP-2000;	2000US-0235484P.
PR	27-SEP-2000;	2000US-0235834P.
PR	27-SEP-2000;	2000US-0235836P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236368P.
PR	29-SEP-2000;	2000US-0236369P.
PR	29-SEP-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0236802P.
PR	02-OCT-2000;	2000US-0237037P.
PR	02-OCT-2000;	2000US-0237038P.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 19:28:36 ; Search time 6476.91 Seconds
(without alignments)
11222.365 Million cell updates/sec

Title: US-10-078-090-48
Perfect score: 1677
Sequence: 1 gaggttgcggcggtgccaaggc.....aagcgagaccaggggggaga 1677

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	836.4	49.9	132292	9	AL138895	AL138895 Human DNA
C 2	829.4	49.5	171627	2	AF336382	AF336382 Homo sapi
C 3	742.4	44.3	3566	9	AB040959	AB040959 Homo sapi
C 4	610.2	36.4	3178	10	AK122523	AK122523 Mus muscu
C 5	585	34.9	2763	10	AY227205	AY227205 Rattus no
C 6	554.8	33.1	2902	9	HSM800885	AL110228 Homo sapi
C 7	420.2	25.1	1718	6	AX714098	AX714098 Sequence
C 8	420.2	25.1	1718	9	AK056190	AK056190 Homo sapi
C 9	180	10.7	237260	2	AC131430	AC131430 Rattus no
C 10	180	10.7	333531	2	AC127863	AC127863 Rattus no
C 11	172.6	10.3	214370	10	AL683828	AL683828 Mouse DNA
C 12	171.8	10.2	219809	2	AF336379	AF336379 Mus muscu
C 13	108.6	6.5	16325	2	AC015144	AC015144 Drosophil
C 14	108.6	6.5	63934	2	AC010699	AC010699 Drosophil
C 15	108.6	6.5	110000	3	AE003536_1	Continuation (2 of
C 16	108.6	6.5	173509	3	AC010031	AC010031 Drosophil
C 17	108.6	6.5	216041	2	AC112524	AC112524 Drosophil
C 18	107.6	6.4	1805	9	BC029054	BC029054 Homo sapi
C 19	107.6	6.4	2037	9	AK026862	AK026862 Homo sapi
C 20	96.2	5.7	179862	2	BX470066	BX470066 Danio rer
C 21	96.2	5.7	191463	2	BX510367	BX510367 Danio rer
C 22	96.2	5.7	219657	2	BX470198	BX470198 Danio rer
C 23	94.6	5.6	191463	2	BX510367	BX510367 Danio rer
C 24	89.6	5.3	2093	5	BC051776	BC051776 Danio rer
C 25	87.8	5.2	139973	2	AC135372	AC135372 Rattus no
C 26	86.6	5.2	211588	2	BX537348	BX537348 Danio rer
C 27	86.2	5.1	46643	10	AL732512	AL732512 Mouse DNA
C 28	86.2	5.1	214625	10	AC116582	AC116582 Mus muscu
C 29	85.2	5.1	166893	9	AP002987	AP002987 Homo sapi
C 30	85.2	5.1	179355	10	AC115631	AC115631 Mus muscu
C 31	85.2	5.1	213715	2	BX649528	BX649528 Danio rer
C 32	85	5.1	130176	2	BX005295	BX005295 Danio rer
C 33	85	5.1	203271	2	BX322566	BX322566 Danio rer
C 34	85	5.1	215778	2	BX530074	BX530074 Danio rer
C 35	84.6	5.0	117886	2	AC139385	AC139385 Rattus no
C 36	84.6	5.0	160718	2	BX664714	BX664714 Danio rer
C 37	84.6	5.0	202878	2	BX511101	BX511101 Danio rer
C 38	84.6	5.0	228962	2	AC131587	AC131587 Mus muscu
C 39	84.6	5.0	229574	2	AC123856	AC123856 Mus muscu
C 40	84	5.0	169095	5	BX088685	BX088685 Zebrafish
C 41	84	5.0	230877	5	BX248500	BX248500 Zebrafish
C 42	84	5.0	237343	2	BX569792	BX569792 Danio rer
C 43	83.8	5.0	165993	5	AL772362	AL772362 Zebrafish
C 44	83.6	5.0	135904	2	AC139157	AC139157 Mus muscu
C 45	83.6	5.0	142803	2	AC013771	AC013771 Homo sapi

ALIGNMENTS

RESULT 1
AL138895/c
LOCUS AL138895 132292 bp DNA linear PRI 27-JUN-2001
DEFINITION Human DNA sequence from clone RP11-9M16 on chromosome 9, complete sequence.
ACCESSION AL138895
VERSION AL138895.16 GI:14575083
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 132292)
AUTHORS Skuce,C.
TITLE Direct Submission

JOURNAL Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Jun 28, 2001 this sequence version replaced gi:14456172. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-9M16 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-9M16 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-9M16 is at 132292 in this sequence. The true left end of clone RP11-402G3 is at 124695 in this sequence. The true right end of clone RP11-8211 is at 2000 in this sequence.

FEATURES

Source	Location/Qualifiers
1. .132292	
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="9"	
/clone="RP11-9M16"	
/clone_lib="RPCI-11.1"	
repeat_region 1. .32	
/note="AluY repeat: matches 1. .32 of consensus"	
repeat_region 74. .301	
/note="L1MC1 repeat: matches 5718. .5955 of consensus"	
repeat_region 399. .696	
/note="AluSq repeat: matches 1. .298 of consensus"	
repeat_region 1186. .1273	
/note="MIR repeat: matches 49. .142 of consensus"	
repeat_region 2271. .2435	
/note="MER63A repeat: matches 46. .206 of consensus"	
repeat_region 2467. .2553	
/note="MIR repeat: matches 79. .165 of consensus"	
repeat_region 2587. .2713	
/note="L2 repeat: matches 2622. .2740 of consensus"	
repeat_region 2911. .3207	
/note="AluSc repeat: matches 5. .301 of consensus"	
repeat_region 3214. .3262	
/note="MIR repeat: matches 115. .166 of consensus"	
repeat_region 4193. .4262	
/note="MIR repeat: matches 60. .130 of consensus"	
misc_feature 4368. .5308	
/note="CpG island"	
/evidence=not_experimental	
repeat_region 5299. .5606	
/note="AluX repeat: matches 12. .312 of consensus"	
repeat_region 5686. .5787	
/note="MIR repeat: matches 141. .262 of consensus"	
repeat_region 6714. .6799	

repeat_region 6995. .7437

repeat_region 7440. .7502

repeat_region 8034. .8179

repeat_region 8160. .8271

repeat_region 10037. .10131

repeat_region 10604. .10711

repeat_region 11151. .11327

repeat_region 11579. .11883

repeat_region 11886. .12024

repeat_region 12598. .12717

repeat_region 13839. .14002

repeat_region 14071. .14205

repeat_region 15216. .15653

repeat_region 15702. .15995

repeat_region 16113. .16847

repeat_region 16875. .16922

repeat_region 17482. .17573

repeat_region 18163. .18379

repeat_region 18972. .19080

repeat_region 20542. .20994

repeat_region 21421. .21521

repeat_region 21523. .21758

repeat_region 22863. .23243

repeat_region 23319. .23495

repeat_region 24438. .24542

repeat_region 24993. .25526

repeat_region 25630. .25867

repeat_region 25889. .26233

repeat_region 26257. .26381

repeat_region 26919. .27000

repeat_region 27292. .27435

repeat_region 29060. .29358

repeat_region 29523. .29664

repeat_region 29669. .29777

repeat_region 30134. .30243

repeat_region 30460. .30627

/note="MIR repeat: matches 20. .114 of consensus"

/note="MLT1D repeat: matches 3. .503 of consensus"

/note="MIR repeat: matches 82. .144 of consensus"

/note="LTR16A repeat: matches 200. .354 of consensus"

/note="LTR16C repeat: matches 255. .368 of consensus"

/note="MIR repeat: matches 14. .123 of consensus"

/note="L2 repeat: matches 2405. .2521 of consensus"

/note="MIR repeat: matches 18. .205 of consensus"

/note="AluX repeat: matches 1. .304 of consensus"

/note="MIR repeat: matches 97. .243 of consensus"

/note="MIR repeat: matches 52. .181 of consensus"

/note="MIR repeat: matches 80. .258 of consensus"

/note="MIR repeat: matches 92. .226 of consensus"

/note="MLT2CB repeat: matches 1. .492 of consensus"

/note="L2 repeat: matches 2372. .2688 of consensus"

/note="L2 repeat: matches 1940. .2707 of consensus"

/note="12 copies 4 mer atga 81% conserved"

/note="MIR repeat: matches 125. .230 of consensus"

/note="MIR repeat: matches 23. .241 of consensus"

/note="MIR repeat: matches 34. .144 of consensus"

/note="Charlie4 repeat: matches 26. .506 of consensus"

/note="L2 repeat: matches 2045. .2158 of consensus"

/note="AluJo repeat: matches 1. .232 of consensus"

/note="MSTA repeat: matches 1. .426 of consensus"

/note="MIR repeat: matches 5. .192 of consensus"

/note="MIR repeat: matches 28. .140 of consensus"

/note="L2 repeat: matches 2099. .2709 of consensus"

/note="MIR repeat: matches 16. .262 of consensus"

/note="L2 repeat: matches 2363. .2674 of consensus"

/note="MIR repeat: matches 28. .153 of consensus"

/note="MIR repeat: matches 64. .147 of consensus"

/note="MIR repeat: matches 102. .254 of consensus"

/note="AluX repeat: matches 1. .298 of consensus"

/note="L2 repeat: matches 2593. .2734 of consensus"

/note="MIR repeat: matches 17. .140 of consensus"

/note="L2 repeat: matches 2167. .2289 of consensus"

/note="MER5B repeat: matches 1. .173 of consensus"

repeat_region	30658. .30770	/note="L2 repeat: matches 2572. .2690 of consensus"
repeat_region	31123. .31357	/note="L2 repeat: matches 2514. .2731 of consensus"
repeat_region	32095. .32170	/note="19 copies 4 mer acac 88% conserved"
repeat_region	32365. .32396	/note="16 copies 2 mer tc 84% conserved"
repeat_region	32426. .32485	/note="MIR repeat: matches 96. .152 of consensus"
repeat_region	33041. .33127	/note="MIR repeat: matches 20. .115 of consensus"
repeat_region	33494. .33764	/note="AluJo repeat: matches 4. .281 of consensus"
repeat_region	34352. .34417	/note="L2 repeat: matches 2641. .2709 of consensus"
repeat_region	34554. .34848	/note="AluSx repeat: matches 1. .293 of consensus"
repeat_region	34946. .35142	/note="MIR repeat: matches 8. .207 of consensus"
repeat_region	35436. .35738	/note="AluSx repeat: matches 1. .303 of consensus"
repeat_region	35834. .35965	/note="MIR repeat: matches 2. .147 of consensus"
repeat_region	36769. .36954	/note="MER53 repeat: matches 1. .188 of consensus"
repeat_region	37658. .37805	/note="MIR repeat: matches 40. .190 of consensus"
repeat_region	37865. .37962	/note="MIR repeat: matches 49. .140 of consensus"
repeat_region	39025. .39247	/note="MIR repeat: matches 14. .252 of consensus"
repeat_region	39184. .39263	/note="L2 repeat: matches 2671. .2750 of consensus"
repeat_region	39516. .39569	/note="MIR repeat: matches 98. .154 of consensus"
repeat_region	40503. .40600	/note="MIR repeat: matches 30. .126 of consensus"
repeat_region	40730. .40829	/note="MIR repeat: matches 46. .146 of consensus"
repeat_region	40854. .40893	
Query Match 49.9%; Score 836.4; DB 9; Length 132292;		
Best Local Similarity 95.5%; Pred. No. 2.7e-183;		
Matches 861; Conservative 0; Mismatches 41; Indels 0; Gaps 0;		
QY	539	AGGTTGGGACACAGATTCTAGAACTGAATGGGGGAGCTTTCTCAACATCCTACACGACG 598
Db	32923	AGGTTGGGACACAGATTCTAGAACTGAATGGGGGAGCTTTCTCAACATCCTACACGACG 32864
QY	599	AGGCTGTACAGGTGCTTAAGTCATCTCGCACCTCATCTGACAGTGAAGGACGTCGGGA 658
Db	32863	AGGCTGTACAGGTGCTTAAGTCATCTCGCACCTCATCTGACAGTGAAGGACGTCGGGA 32804
QY	659	GGCTGCCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTCCCGGATCA 718
Db	32803	GGCTGCCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTCCCGGATCA 32744
QY	719	GGGAGACCATGGCGAACTCGGCAGGGTCTGGCCACTCTGCTCGCTCCAATCTCCAGACCC 778
Db	32743	GGGAGACCATGGCGAACTCGGCAGGGTCTGGCCACTCTGCTCGCTCCAATCTCCAGACCC 32684
QY	779	CAGGGCCATTCTGAAAGCCAGTGATAGTGCCTCCCATCCCTCCACCGCCCTGGCTCTC 838
Db	32683	CAGGGCCATTCTGAAAGCCAGTGATAGTGCCTCCCATCCCTCCACCGCCCTGGCTCTC 32624
QY	839	CTCTCAGCCTGCAGTCCCCACACACAGGGCCCTCCATTTGGCAGGACATGACCTGGGCACAT 898
Db	32623	CTCTCAGCCTGCAGTCCCCACACACAGGGCCCTCCATTTGGCAGGACATGACCTGGGCACAT 32564
QY	899	CCCTCTCCTCTCTTGGCCCTCAGTTTCCCATGGAAGCTGAAATACACCATCCAACTGTC 958
Db	32563	CCCTCTCCTCTCTTGGCCCTCAGTTTCCCATGGAAGCTGAAATACACCATCCAACTGTC 32504

QY	959	TCATTCTTTATTGTCCCAAAATTACTTAACCTCATTTCTATAGACCTTAGTTGCTTCATCC 1018
Db	32503	TCATTCTTTATTGTCCCAAAATTACTTAACCTCATTTCTATAGACCTTAGTTGCTTCATCC 32444
QY	1019	AAAAAGTGGGACCATAAACCCCTGCCCTCATCCAGATCTGTGCAGATGAAAAGAGAGGGAG 1078
Db	32443	AAAAAGTGGGACCATAAACCCCTGCCCTCATCCAGATCTGTGCAGATGAAAAGAGAGGGAG 32384
QY	1079	GGAGAGGGAAGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAGGCCACCAGGCTGGATC 1138
Db	32383	GGAGAGGGAAGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAGGCCACCAGGCTGGATC 32324
QY	1139	CCATGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCATAAACTGAGATCCAGTTGCCAGGT 1198
Db	32323	CCATGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCATAAACTGAGATCCAGTTGCCAGGT 32264
QY	1199	GGCTGCATAGTTGCCAACAGTGTAAATGTGTACCTTTTGATCTTTCATCAGAAAATCTCAGG 1258
Db	32263	GGCTGCATAGTTGCCAACAGTGTAAATGTGTACCTTTTGATCTTTCATCAGAAAATCTCAGG 32204
QY	1259	CTGGTGGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTGTCT 1318
Db	32203	CTGGTGGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTGTCT 32144
QY	1319	TCTCTGTCTCCT 1378
Db	32143	TCTCTGTCTCCT 32084
QY	1379	CACCTCTTTCATCTATCATACATAGTAGTATAATAATAATATTAGAGAGATACACAGA 1438
Db	32083	CCTCCTTTCACCTGCATAGTATAATAACATTTTCAAAGTTACTTGCCAAACATCTAGAAAGA 32024
QY	1439	AA 1440
Db	32023	TA 32022

RESULT 2	AF336382/c	171627 bp	DNA	linear	HTG 02-APR-2001
LOCUS	AF336382	Homo sapiens chromosome 9 clone R-3183P23 map 9q33, *** SEQUENCING			
DEFINITION	IN PROGRESS ***	3 unordered pieces.			
ACCESSION	AF336382				
VERSION	AF336382.1	GI:13507299			
KEYWORDS	HTG; HTGS_PHASE1.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 171627)				
AUTHORS	Rump,A., Varela,A., Mburu,P., Brown,S.D.M. and Rosenthal,A.				
TITLE	Mouse chromosome 4 genomic sequence				
JOURNAL	Unpublished				
AUTHORS	2 (bases 1 to 171627)				
TITLE	Rump,A.				
JOURNAL	Direct Submission				
COMMENT	Submitted (17-JAN-2001) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany				
	* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces				
	* is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as				
	* runs of N, but the exact sizes of the gaps are unknown.				
	* This record will be updated with the finished sequence as soon as it is available and the accession number will				
	* be preserved.				
	* 1 803: contig of 803 bp in length				
	* 804 903: gap of unknown length				
	* 904 8454: contig of 7551 bp in length				
	* 8455 8554: gap of unknown length				
	* 8555 171627: contig of 163073 bp in length.				
FEATURES	Location/Qualifiers				


```

source      1. .171627
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q33"
/clone="R-3183P23"

ORIGIN
Query Match      49.5%; Score 829.4; DB 2; Length 171627;
Best Local Similarity 94.0%; Pred. No. 1.2e-181;
Matches 888; Conservative 0; Mismatches 46; Indels 11; Gaps 2;

QY 539 AGGTTGGGACACAGATCTAGAAAGTGAATGGCGGAGCTTTCTCAACATCTACACGACG 598
Db 101687 AGGTTGGGACACAGATCTAGAAAGTGAATGGCGGAGCTTTCTCAACATCTACACGACG 101628

QY 599 AGGCTGTGAGGCTGTTAAGTCACTCGGCACCTCATCTGACAGTGAGGACGTCGGGA 658
Db 101627 AGGCTGTGAGGCTGTTAAGTCACTCGGCACCTCATCTGACAGTGAGGACGTCGGGA 101568

QY 659 GGTGCCCCCATGCCCGACCACTGTGGACGAGACCAAGTGATCGCCAGTTCCTCGGATCA 718
Db 101567 GGTGCCCCCGTGCCCGACCACTGTGGACGAGACCAAGTGATCGCCAGTTCCTCGGATCA 101508

QY 719 GGGAGACCATGGCGAACTCGGCAGGGTCTGGCCACTCTGTCGCTCCAATCTCAGACCC 778
Db 101507 GGGAGACCATGGCGAACTCGGCAGGGTCTGGCCACTCTGTCGCTCCAATCTCAGACCC 101448

QY 779 CAGGCCCATTCTGAAAGCCAGTGATAGCTGCTCCCATCCCTCCACCGCCCTGGCTCTC 838
Db 101447 CAGGCCCATTCTGAAAGCCAGTGATAGCTGCTCCCATCCCTCCACCGCCCTGGCTCTC 101388

QY 839 CTCTCAGCCTGCAGTCCCAACACACGAGGCCCTCCATTGGCAGACATGACCTGGGCACAT 898
Db 101387 CTCTCAGCCTGCAGTCCCAACACACGAGGCCCTCCATTGGCAGACATGACCTGGGCACAT 101328

QY 899 CCCTCTCCTCTCTTGGCCTCAGTTTCCCCATGGAAAGCTGAAATACACCATCCAACTGTC 958
Db 101327 CCCTCTCCTCTCTTGGCCTCAGTTTCCCCATGGAAAGCTGAAATACACCATCCAACTGTC 101268

QY 959 TCATTCTTTATTGTGCCCAAAATTACTTAACCTATCTATAGACCTTAGTTGCTTCATCC 1018
Db 101267 TCATTCTTTATTGTGCCCAAAATTACTTAACCTATCTATAGACCTTAGTTGCTTCATCC 101208

QY 1019 AAAAGTGGGACCATAACCTGCCCCTCATCCAGATC---TGTGCAGATGAAAGAGAG 1074
Db 101207 AAAAGTGGGACCATAACCTGCCCCTCATCCAGATCTGTGTGTCAGATGAAAGAGAG 101148

QY 1075 GGAGGGAGGGAAAGAGAGAGATGCTTTGGGGTGTATTGTGGCCAGAGGCCACCAGGCTG 1134
Db 101147 GGAGGGAGGGAAAGAGAGAGATGCTTTGGGGTGTATTGTGGCCAGAGGCCACCAGGCTG 101088

QY 1135 GATCCCATGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCATAAACTGAGATCCAGTTGCC 1194
Db 101087 GATCCCATGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCATAAACTGAGATCCAGTTGCC 101028

QY 1195 AGGTGGCTGCATAGTTGCCAACAGTGTAAATGTGTACCTTTTGTATCTTTCATCAGAAATCT 1254
Db 101027 AGGTGGCTGCATAGTTGCCAACAGTGTAAATGTGTACCTTTTGTATCTTTCATCAGAAATCT 100968

QY 1255 CAGGCTGGTGGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTG 1314
Db 100967 CAGGCTGGTGGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTG 100908

QY 1315 TGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1374
Db 100907 TGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 100848

QY 1375 TCCTCACTCTTTTCATCCTATCAATTACATAGTAGTATAATAATAATATTAGAGAGATACA 1434
Db 100847 TGTCCCCCTTCCCTTCTTC-----ACCTGCATAGTATTAAACATTTTCAAAGTTAC 100798

```

QY 1435 CAGAAAATATATAGAGAAGATACAGTGTCTCTATAAAAAAAA 1479

Ddb 100794 TTGCCAACATCTAGAAAGATACCAGGTTTCTATAAAAAAAA 100750

RESULT 3

AB040959

LOCUS 3566 bp mRNA linear PRI 22-FEB-2001

DEFINITION Homo sapiens mRNA for KIAA1526 protein, partial cds.

ACCESSION AB040959

VERSION AB040959.1 GI:7959318

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (sites)

AUTHORS Nagase,T., Kikuno,R., Ishikawa,K., Hirosawa,M. and Ohara,O.

TITLE Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro

JOURNAL DNA Res. 7 (2), 143-150 (2000)

MEDLINE 20277482

PUBMED 10819331

REFERENCE 2 (bases 1 to 3566)

AUTHORS Chara,O., Nagase,T. and Kikuno,R.

TITLE Direct Submission

JOURNAL Submitted (04-APR-2000) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913, Fax:81-438-52-3914)

FEATURES

source Location/Qualifiers

1..3566

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="fj04743"

/clone_lib="pBluescriptII SK plus"

1..3566

/gene="KIAA1526"

<1..2893

/gene="KIAA1526"

/note="Start codon is not identified. fj04743 cDNA clone for KIAA1526 has a 1196-bp insertion after the position 1335 of the sequence of KIAA1526."

/codon_start=2

/product="KIAA1526 protein"

/protein_id="BAA96050.1"

/db_xref="GI:7959319"

/translation="SPRPRIQLYPCSPNPGSSRRRLRAPGDPVAVPVPARTAAPPTP RALRARPLGVEMKAPLDGLSVSSSTGLSAGAGGGGAGRLLSANVRQLHQDL TALLSEAREQFTHCLNAYHARRNVFDLRTLEVLDSVKKRLLPMLRLVTPRSDQL LFDQYTAEGLYLPATTPYQPAGWGPDSAGPEVRLVSLRRAKAHEGLGFSIRGSEH GVGIYVSLVEPGSLAEKGLRVGDQILRVNDKSLARVTHAEAVKALXGSKLVLSVYS AGRIPGVYTNHIYVDPQGRSISPPSGLPQPHGALRQQEGRDRRTLHLLQGQDEK KNLVLGDGRSLGLTIRGAEGYGLGIYITGVDPGSEAGSGGLKVGDDQILEVNGRSLN ILHDEAVRLKSSRHILTVKDVGRPLHARTVTDTKWIASRIRRETMANSAAGFLGDL TTEGINKPGFYKGPAGSQVTLSSLGNTQTRVLLEEQARHLLNEQEHATMAYVLDEYRG SVSVEALVMALFKLNLTHAKFSLSEVRGTISPDLERFDHLVLRREIESMKARQPPG PAGDITYSMVSYSDTGSSTGSHGTSTTVSSARNTLDLEETGEAVQGNINALPDVSVDD VRSTSQGLSSFPLPRPPLAQGNDPLPQPKRLGREDLPSPSPMSPSGTGFVSAPQN RSPAGTAPTPTGTSSAQDLPSPIYASVSPANPSSKRLDHALLVNQHPHIGPFPRVQ SPPLKPSAEATVAGGCLLPSPSGHPDQTGTHQFVMEVHRPDSEPDVNEVRALP QTRASTLSQLSDSGQTLSDESDGVDAEAEASAPGRQSVSTKSRSKELPRNERPT DGANKPGLLEPTSTLVRVKKSAATGLGIAEGGANTRQPLPRIVTIQRGGSANCCGL KVGHVILEVNGLTLRGKHEHREARIIEAFKTKDRDYIDFLVTEFNVL"

gene

CDS

ORIGIN

Query Match 44.3%; Score 742.4; DB 9; Length 3566;

Best Local Similarity 99.9%; Pred. No. 1e-161;

Matches 743; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGGGGCGTGCCAAAGGCCACAGGGCTTGGGCTTCCAGCATCCGTGGGGCTCGGAGCAC 63
Db 596 TTGGGGCGTGCCAAAGGCCACAGGGCTTGGGCTTCCAGCATCCGTGGGGCTCGGAGCAC 655
QY 64 GGCGTGGGCATCTACGTGTCTCTGGTGGAAACAGGCTCTCTAGCTGAGAGGAGACTG 123
Db 656 GGCGTGGGCATCTACGTGTCTCTGGTGGAAACAGGCTCTCTAGCTGAGAGGAGACTG 715
QY 124 CGGTGCGGGACACAGATTCTGGCGCTCAACGACAAATCCCTGGCCCGGTGACCCACGG 183
Db 716 CGGTGCGGGACACAGATTCTGGCGCTCAACGACAAATCCCTGGCCCGGTGACCCACGG 775
QY 184 GAGCGCGTCAAGGCTCTGAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGG 243
Db 776 GAGCGCGTCAAGGCTCTGAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGG 835
QY 244 CGATCCCTGGGGCTACGTACCAACCAACATCTACACCTGGTGGACCCCGAGGGCCGC 303
Db 836 CGATCCCTGGGGCTACGTACCAACCAACATCTACACCTGGTGGACCCCGAGGGCCGC 895
QY 304 AGCATCTCCCCACCCCTGGGCTGCCCCAGCCCCAGCTGGTGGTGGTGGTGGTGGTGG 363
Db 896 AGCATCTCCCCACCCCTGGGCTGCCCCAGCCCCAGCTGGTGGTGGTGGTGGTGGTGG 955
QY 364 GGTACCGGAGGACCCCTGACCTCTCTGCAAGAGGGGATGAGAAAGGTGAACCTG 423
Db 956 GGTACCGGAGGACCCCTGACCTCTCTGCAAGAGGGGATGAGAAAGGTGAACCTG 1015
QY 424 GTGCTGGGGGACGGCGGCTCCCTGGGCTCACGATCCGTGGGGAGTGAGTACGGCTT 483
Db 1016 GTGCTGGGGGACGGCGGCTCCCTGGGCTCACGATCCGTGGGGAGTGAGTACGGCTT 1075
QY 484 GGCATTTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGCGGCTCAAGGTT 543
Db 1076 GGCATTTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGCGGCTCAAGGTT 1135
QY 544 GGGACACAGATTCTAGAGTGAATGGCGGAGCTTCTCAACATCTCAACAGCAGGCT 603
Db 1136 GGGACACAGATTCTAGAGTGAATGGCGGAGCTTCTCAACATCTCAACAGCAGGCT 1195
QY 604 GTCAGGCTGCTTAAGTCACTCGGCACCTCATCTGACAGTGAAGAGCTCGGGAGGCTG 663
Db 1196 GTCAGGCTGCTTAAGTCACTCGGCACCTCATCTGACAGTGAAGAGCTCGGGAGGCTG 1255
QY 664 CCCATGCCCGCACCATCTGGACGAGACCAAGTGGATCGCCAGTTCCTCCGGATCAGGGAG 723
Db 1256 CCCATGCCCGCACCATCTGGACGAGACCAAGTGGATCGCCAGTTCCTCCGGATCAGGGAG 1315
QY 724 ACCATGGCGAACTCGGAGGGTCT 747
Db 1316 ACCATGGCGAACTCGGAGGGTTT 1339

RESULT 4
AK122523
LOCUS AK122523 3178 bp mRNA linear ROD 15-MAR-2003
DEFINITION Mus musculus mRNA for mKIAA1526 protein.
ACCESSION AK122523
VERSION AK122523.1 GI:28972777
KEYWORDS FLI_CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Okazaki,N., Kikuno,R., Ohara,T., Inamoto,S., Aizawa,H., Yuasa,S.,
Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA
gene: II. The complete nucleotide sequences of 400 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
JOURNAL DNA Res. 10, 35-48 (2003)
REFERENCE 2 (bases 1 to 3178)

AUTHORS Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing: Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.

FEATURES
Location/Qualifiers
1..3178
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbh03417"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"

gene
1..3178
/gene="mKIAA1526"
CDS
1..2610
/gene="mKIAA1526"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence=not experimental
/product="mKIAA1526 protein"
/protein_id="BAC65805.1"
/db_xref="GI:28972778"

/translation="ANVRQLHQALTALLSEPEREQFTHCLNAYHARRNVFDLVRTLRV
LLDSPVKRRLLPMLRLVPRSDQLFDQYTAELGLYPATTPYRQPAWAAPDAGPGEV
RLVSLRRAKAHEGLGFSIRGGSEHGVIYVSLVEGLAEKEGLRVDQILRVNDKSL
ARVTHAEAVKALKSKKLVLVSVYAGRIYVSVNHIYTWDPQGRSTSPSSLPQPH
GSTLRQREDDRRSTLHLQSGDEKKNLVLDGRSLGLTIRGAEYGLGIYITGVDPG
SEAESGLKVGDOILEVNGRSFLNHLDEAVKLKSSRHLLTVKDVGRLLPHARTTVD
QTKWIASRRIGESVANSAGPPGDHTEGSKPGFKYKPGAGSVQTVLSSLGNQTRALLDD
QARHLLTEQERATMYLAQYRGGTISVEAMVMALFELLNTHAKFSILSEVRSIIISPO
DLDRFDHLVLRREIESMKARQPPGPGVDYISMVSYSDTSGSSGSHGTSTTVSSARNT
LDLEGTGETTGGSTNALPDVSDVDKSPEDLPKIPKPPPPPLAOGHDLRLGQPRKP
GREDPAPLSAAHSGIVFAPNRSPPPGPTAPTPGSSAQDPSPIYASISHANPSS
RKPLDTHLALVNOHPGTPFPVQSPHKLSPPAETPGAGACLPSPSPSEHPDVGANO
HFVLVEHRPDSEPDVNEVRALPQRTSTLSQSDSGQTLSEDSGVDAGETASTGR
GRQTASAKNKGKQPTERTAEAGANKPGLLEPTSTLVRKSAATLGLIAEGGANT
RQPLPRIVTIQGGSAHNCGQLKVGHVILEVNGQTLRGKEHKEAARIIEAFKTKERD
YIDPLVTEFNVL"

ORIGIN
Query Match 36.4%; Score 610.2; DB 10; Length 3178;
Best Local Similarity 88.8%; Pred. No. 4.9e-131;
Matches 660; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 5 TGCGGCGTGCCAAAGGCCACAGGGCTTGGGCTTCCAGCATCCGTGGGGCTCGGAGCACG 64
Db 320 TGCGGCGTGCCAAAGGCCACAGGGCTTGGGCTTCCAGCATCCGTGGGGCTCGGAGCACG 379
QY 65 GCGTGGGCATCTACGTGTCTCTGGTGGAAACAGGCTCTCTAGCTGAGAGGAGACTGC 124
Db 380 GCGTGGGCATCTACGTGTCTCTAGTGGAGCCCGGCTCCCTGGCAGAGAGGAGGTTGC 439
QY 125 GGGTCGGGACACAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGTGACCCACGGCG 184
Db 440 GGGTCGGGACACAGATTCTGCGCGTCAACGATAAATCTCTAGCCCGGTGACCCACGGCG 499
QY 185 AGGCGGTCAAGGCTCTGAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGGC 244
Db 500 AGGCTGTCAAGGCTCTCAAGAGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGGC 559
QY 245 GCATCCCTGGGGCTACGTACCAACACACATCTACACCTGGTGGACCCGAGGGCCGCA 304
Db 560 GTATCCAGGGGGCTATGTGACCAACACACATCTACACCTGGTGGACCCGAGGGTCGAA 619
QY 305 GCATCTCCCCACCCCTCGGGCGCTGCCCCAGCCCCACCGTGTGTCCTTGAGGAGCAGGAGG 364

Db 620 GCACATCCCTCCCTCAGCCTGCCCCAGCCCATGGCAGCACCCCTGAGACAGCGTGAAG 679
QY 365 GTGACCGGAGGAGCACCTCGCACCTCTCTGCAAGAGGGGATGAGAAAAGGTGAACCTGG 424
Db 680 ATGACCGAAGGAGTACCTCCACCTCTCGAGAGTGGAGATGAGAAAAGGTGAACCTGG 739
QY 425 TGCTGGGGGACGGCCGCTCCCTGGCCTCACGATCCGTGGGGAGCTGAGTACGGCCTTG 484
Db 740 TGTGGGGGACGGCCGCTCTTGGCCTCACGATCCGAGGTGGAGCAGAGTACGGCCTTG 799
QY 485 GCATTTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAAGGCAGCGGGCTCAAGGTTG 544
Db 800 GCATTTACATCACTGGTGTGGACCCAGGCTCTGAAGCAGAGAGCAGCGGGCTCAAGGTTG 859
QY 545 GGGACCAAGTTCTAGAAAGTGAATGGCGGAGCTTTCTCAACATCCTACACAGCAGAGGCTG 604
Db 860 GAGACCAGATTCTGGAGGTGAATGGCGGAGCTTTCTCAACATCCTGCATGATGAGGCAG 919
QY 605 TCAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGC 664
Db 920 TGAAGCTGCTCAAGTCATCCCGGCACCTCATCTGACGGTGAAGGACGTCGGAAGGCTGC 979
QY 665 CCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCCCGGATCGGGGAAA 1039
Db 980 CCCACGCACGTACCAACCGTGGACCAAGTGGATCGCCAGTTCCCGGATCGGGGAAA 1039
QY 725 CCATGCGAACTCGGCAGGGTCT 747
Db 1040 GCGTCGCCAACTCAGCAGGGTTT 1062

RESULT 5
AY227205
LOCUS
DEFINITION
AY227205
Rattus norvegicus CASK-interacting protein CIP98 (Cip98) mRNA, complete cds.
AY227205
AY227205.1 GI:29373052

SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS
Yap,C.C., Liang,F., Yamazaki,Y., Muto,Y., Kishida,H., Hayashida,T., Hashikawa,T. and Yano,R.
TITLE
CIP98, a novel PDZ domain protein, is expressed in the central nervous system and interacts with calmodulin-dependent serine kinase

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
J. Neurochem. 85 (1), 123-134 (2003)
22529415
12641734
2 (bases 1 to 2763)
Yap,C.C., Liang,F., Yamazaki,Y., Muto,Y., Kishida,H., Hayashida,T., Hashikawa,T. and Yano,R.

TITLE
JOURNAL
Submitted (29-JAN-2003) Brain Science Institute (BSI), RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

FEATURES
source
1. .2763
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="WST"
/db_xref="taxon:10116"
1. .2763
/gene="Cip98"

gene
CDS
1. .2763
/gene="Cip98"
/note="98 kDa"
/codon_start=1
/product="CASK-interacting protein CIP98"

/protein_id="AA072534.1"
/db_xref="GI:29373053"
/translation="MNAQLDGLSVSSSTGSLGSAAAAAGGGAGLRLLSANVRQLH QALTALLSEPEREQTHCLNAYHARNVFDLVRTLRVLDSPVKRRLLPMLRLVIPRS DQLLFDQYTAEGLYLPATTPYRQPAWAAPDGAGPEVRLVRAKAHEGLGFSIRGG SEHGVIYVSLVEPGSLAEKEGLRVGDQILRVNDKSLARVTHAEAVKALKGSKLVLVS VYSGRIPGGYVTHIYTWVDPQGRSTPPSSPLSHGSTLRHEDDRRSALHLLQSGDE KKNVLVGDGRSLGTLIRGAEYGLIYITGVDPGSEASSGLKVGDQILEVNGRSEL SILHDEAVKLKSSRHLLITVDVGRPLPHARTTVDQTKWIASRIGESITNSAGFPGD LTEEGNKPGFYKGPAGSQVTLSSLNQTRALLDDQARHLLTGEERATMMYYLDQYRG GTISVEALVMALFELLNTHAKFSLSEVRGIIISPDLDLRFDLVLRREIESMKARQPP GPGVDTYSMVYSYSDTGSSTGSHGTSTTVSSARERLLWLIDLMENTLDLEGTCTTQG STNALPDVSDVDVRSPSEDLPGIKPPPPPPPLAQGHRIILGQTRKPVREDSSAPLSSAA HSGIVFSAPRNRSPPPPGIAFTPTGPPSSARDSPPSIYASISHANPSSRKPLDTHL ALVNOHPIGPPRVSPPHLKSPPAEAPGAGACLPSPSPSEHADAMGANQHFLVLEVH RPDSEPDVNEURALPQTRTASTLSQLSDSGQTLSEDSGVDAGEIEASTSGRGRQTANT KNKNGKELPOTERTTEGANKPPGLLEPTSTLIRVKSAAATLGLAIEGGANTRQPLPRI VTIORGGSAHNCGQLKVGHVILEVNGQTLRGKEHREARVIAEAFKTKERDYIDFLVT EFNVM"

ORIGIN

Query Match 34.9%; Score 585; DB 10; Length 2763;
Best Local Similarity 87.8%; Pred. No. 3.4e-125;
Matches 652; Conservative 0; Mismatches 85; Indels 6; Gaps 1;

QY 5 TCGGCGGTGCCAAGGCCACGAGGGCTTGGGCTTACGATCCGTGGGGGCTCGGAGCACG 64
Db 431 TCGACGCGCCCAAGGCCACGAGGGCTTGGGCTTACGATCCGCGGGGCTCGGAGCACG 490
QY 65 GCGTGGGCATCTACGTGTCTCTGGTGGAAACGAGCTCTTAGCTGAGAAGGAAGGACTGC 124
Db 491 GCGTGGGCATCTACGTGTCTCTAGTGGAGCCAGGCTCCCTGGCAGAGAAGGAGGTTAC 550
QY 125 GGGTGGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGTGACCCACGGG 184
Db 551 GGGTGGGGACCAAGATTCTGCGCGTCAACGATAAATCTTAGCCCGGTGACCCACGGG 610
QY 185 AGGCCGTCAAAGGCTCTGAAGGGTCCAAAGAGCTGTTGTGTCTGTACTCAGCAGGGC 244
Db 611 AGGCTGTCAAGGCTCTCAAGGGTCCAAAGAGTTGGTGTGTCTGTATCTCAGCTGGGC 670
QY 245 GCATCCCTGGGGGCTACGTACCAACCAACATCTACACCTGGGTGGACCCGCGGCGCA 304
Db 671 GTATACCCGGGGCTATGTACCAACCAACATCTACACCTGGGTGGACCCCGGGTCAAA 730
QY 305 GCATCTCCCAACCCCTCGGGCTTGCCTCAGCCCGCCAGCGTGGTGCCTGAGGCAGGAGG 364
Db 731 GCACATCCCTCCCTCCAGCCT-----TCCCCATGGCAGCACCTTGAGACAGCATGAAG 784
QY 365 GTGACCGGAGGAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGG 424
Db 785 ATGATCGAAGGAGTGCCTACACCTCCTGCAGAGTGGAGATGAGAAAAAAGGTGAACCTGG 844
QY 425 TGCTGGGGGACGGCCGCTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGCCTTG 484
Db 845 TGCTGGGGGACGGCCGCTCCTTGGGCTCACAAATCCGAGGTGGAGCCGAGTACGGCCTTG 904
QY 485 GCATTTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGAAGGCGGGCTCAAGGTTG 544
Db 905 GCATTTACATCACTGGTGTGGACCCGGGCTCTGAAGCCGAAAGCAGCGGCGCTCAAGGTTG 964
QY 545 GGGACCAGATTCTAGAAAGTGAATGGCGGAGGCTTTCTCAACATCCTACACGAGGCTG 604
Db 965 GGGACCAGATTCTGGAGGTGAACGGGCGGAGGCTTTCTCAGCATCCTGCATGACGAGGCAG 1024
QY 605 TCAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGC 664
Db 1025 TGAAGCTGCTCAAGTCATCCCGGCACCTCATCTGACCGTGAAGGACGTCGGGAGGCTGC 1084
QY 665 CCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCCCGGATCAGGGAGA 724
Db 1085 CCCACGCACGTACCAACCGTGGACCAAGTGGATCGCCAGTTCCCGGATCGGGGAAA 1144

Qy 725 CCATGGCGAACTCGGACGGTCT 747 2902 bp mRNA linear PRI 18-FEB-2000
Db 1145 GCATCACCACCTCAGCAGGGTTT 1167
HSM800885
LOCUS HSM800885 2902 bp mRNA linear PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp434N014 (from clone DKFZp434N014);
partial cds.
ACCESSION AL110228
VERSION AL110228.1 GI:5817166
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2902)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-1999) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp434N014) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES
source
1..2902
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp434N014"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2902
/gene="DKFZp434N014"
<1..2173
/gene="DKFZp434N014"
/note="similarity to NY-CO-38 protein"
/codon_start=2
/product="hypothetical protein"
/protein_id="CAB53685.1"
/db_xref="GI:5817167"
/db_xref="GOA:Q9UFZ3"
/db_xref="SPTREMBL:Q9UFZ3"
/translation="KLLGSRGPRLPPCERSVACVQALKGSKLVLSVYSAGRIPGGY
VTNHIYTWDPQGRSISPPSGLPQPHGALRQEGDRRSTLHLQGGDEKKVNLVLGD
GRSLGLTIRGGAEGYGLGYITGVDPGSEAGSLKVGDDQILEVNGRSFNLILHDEAVR
LLKSSRLILTVKDVGLPHARTTVDETKWIASRIRETMANSAGFLDLTTEGINKP
GFYKGPAGSQVTLSSGNQTRVLLEEQARHLNEQEHATMAYLDEYRGGSVSVEALV
MALFKLLNTHAKFSLLEVRGTISPQDLERFDHLVLRREIESMKARPPGPGAGDTYS
MVSYSDTGSSTSGHSTTVSSARNTLDEETCEAVQGNINALPDVSDVDRSTSQL
SSFKPLPRPPLAQNDLPLGQPKLGRDLQFPSSMPCSGTVFSAPQNRSPAGTA
PTPGTSSAQDLPSPIYASVSPANPSSKRPLDAHLALVNQHPGIPFPRVQSPPHLKSP
SABATVAGCLLPSPSGHPDQGTNQHFVMEVHRPDPSEPDVNEVRALPQTRTASTL
SOLSDSGQTLSEDSGVDAEAEASAPGRQSVSTKSRSKELPRNERPTDGANKPPG
LLEPTSTLVRVKKSAATLGLAIEGGANTROPPIVTIQRGSSAHNCCQLKVGHVILE
VNGLTLRGKEHREAARIIAEAFKTKDRDYIDFLVTEFNVL"
2824..2829
/gene="DKFZp434N014"
2846
/gene="DKFZp434N014"
polyA_signal
2824..2829
/gene="DKFZp434N014"
polyA_site
2846
/gene="DKFZp434N014"
ORIGIN
Query Match 33.1%; Score 554.8; DB 9; Length 2902;
Best Local Similarity 99.6%; Pred. No. 3.5e-118;

Matches 556; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 190 GTCAAGGCTCTGAAGGCTCCAAGAAGCTGGTCTGTCTGTACTCAGAGGGCGCATC 249
Db 62 GTCCAGGCTCTGAAGGCTCCAAGAAGCTGGTCTGTCTGTACTCAGAGGGCGCATC 121
Qy 250 CTTGGGGCTACGTACCAACACACATCTACCTGGGTGGACCCCGCAGGGCCGACATC 309
Db 122 CTTGGGGCTACGTACCAACACACATCTACCTGGGTGGACCCCGCAGGGCCGACATC 181
Qy 310 TCCCCACCTCGGGCTGCCCCCAGCCCCACGGTGGTCCCTGAGGCAGCAGAGGGTGAC 369
Db 182 TCCCCACCTCGGGCTGCCCCCAGCCCCACGGTGGTCCCTGAGGCAGCAGAGGGTGAC 241
Qy 370 CGGAGGAGCACCTGACCTCTCTGCAAGGAGGGGATGAGAAAAAGTGAACCTGGTCTG 429
Db 242 CGGAGGAGCACCTGACCTCTCTGCAAGGAGGGGATGAGAAAAAGTGAACCTGGTCTG 301
Qy 430 GGGACGGCCGGTCCCTGGGCTCACGATCCCTGGGGGAGCTGAGTACGGCCTTGGCATT 489
Db 302 GGGACGGCCGGTCCCTGGGCTCACGATCCCTGGGGGAGCTGAGTACGGCCTTGGCATT 361
Qy 490 TACATCACTGGCTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGGAC 549
Db 362 TACATCACTGGCTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGGAC 421
Qy 550 CAGATTCTAGAAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGACGAGGCTGTGAGG 609
Db 422 CAGATTCTAGAAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGACGAGGCTGTGAGG 481
Qy 610 CTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGCCCAT 669
Db 482 CTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGCCCAT 541
Qy 670 GCCCGACCACTGTGGACGAGACCAAGTGGATCGGCAGTTCCTGGATCAGGGAGACCATG 729
Db 542 GCCCGACCACTGTGGACGAGACCAAGTGGATCGGCAGTTCCTGGATCAGGGAGACCATG 601
Qy 730 GCGAACTCGGCAGGGTCT 747
Db 602 GCGAACTCGGCAGGGTCT 619
RESULT 7
AX714098 1718 bp DNA linear PAT 15-APR-2003
LOCUS Sequence 782 from Patent EP1293569.
DEFINITION AX714098
ACCESSION AX714098
VERSION AX714098.1 GI:29889026
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
Masuho,Y.
TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 782 19-MAR-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)
FEATURES Location/Qualifiers
source 1..1718
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 25.1%; Score 420.2; DB 6; Length 1718;
Best Local Similarity 95.0%; Pred. No. 5.6e-87;
Matches 434; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY	4	TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC	63
Db	1076	TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC	1135
QY	64	GGCGTGGGCATCTACGTGCTCTCTGGTGAACCAAGCTCTCTAGCTGAGAAGGAGGACTG	123
Db	1136	GGCGTGGGCATCTACGTGCTCTCTGGTGAACCAAGCTCTCTAGCTGAGAAGGAGGACTG	1195
QY	124	CGGTCGGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG	183
Db	1196	CGGTCGGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG	1255
QY	184	GAGGCCGTCAAGGCTCTGAAGGGCTCCAAGAAGCTGGTGCTGTCTGTACTCAGCAGGG	243
Db	1256	GAGGCCGTCAAGGCTCTGAAGGGCTCCAAGAAGCTGGTGCTGTCTGTACTCAGCAGGG	1315
QY	244	CGCATCCCTGGGGCTACGTCAACCAACCATCTACACCTGGGTGGACCCCGAGGCCCGC	303
Db	1316	CGCATCTCTGGGGCTACGTCAACCAACCATCTACACCTGGGTGGACCCCGAGGCCCGC	1375
QY	304	AGCATCTCCCCACCTCGGGCTGCCCCAGCCCCACGGTGGTGCCCTGAGCAGCAGGAG	363
Db	1376	AGCATCTCCCCACCTCGGGCTGCCCCAGCCCCACGGTGGTGCCCTGAGCAGCAGGAG	1435
QY	424	GTGCTGGGGACGGCCGGTCCCTGGGCGCTCACGATCC	460
Db	1496	GTGGGAAAGGAGGCCAGCCTCTCAGACACCGTATTC	1532
RESULT 8			
AK056190			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			

source		1. 1718	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="NT2R12003344"	
		/cell_line="NT2"	
		/cell_type="teratocarcinoma"	
		/clone_lib="NT2R12"	
		/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction.-majorly NT2 neuron"	
CDS		650..1687	
		/note="unamed protein product"	
		/codon_start=1	
		/protein_id="BAB711117.1"	
		/db_xref="GI:16551527"	
		/translation="MNAPLDGLSVSSSTGSLGSAAGAGGGGAGLRLLSANVRQLHQ	
		ALTALLSEAREBQFTHCLNAYHARNVFDLVRTLRVLLDSPVKRRLPMLRLVPRSD	
		QLLFDQYTAEGDLPATTPYRQPAWGPDGAGPEVRLVSLRRAKAHEGLGFSIRGGS	
		EHGVGIYVSLVEPGLAEKEGLRVGDOIILRVNDKSLARVTHAEAVKALKGSKKLVLVS	
		YSAGRISGGYVTNHIYTWVDPQGRSISPPSGLPQPHGGALRQOEGDRRSTLHLLQGGD	
		EKKVSGVGKGGQPLRHLRILPPNPEQQSCLEAARRGWFCPGSVFPQVCTEGWCFFFAFL	
		FDLCSVCYNTG"	
ORIGIN			
Query Match		25.1%; Score 420.2; DB 9; Length 1718;	
Best Local Similarity		95.0%; Pred. No. 5.6e-87;	
Matches 434; Conservative		0; Mismatches 23; Indels 0; Gaps 0;	
QY	4	TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC	63
Db	1076	TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC	1135
QY	64	GGCGTGGGCATCTACGTGCTCTCTGGTGAACCAAGCTCTCTAGCTGAGAAGGAGGACTG	123
Db	1136	GGCGTGGGCATCTACGTGCTCTCTGGTGAACCAAGCTCTCTAGCTGAGAAGGAGGACTG	1195
QY	124	CGGTCGGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG	183
Db	1196	CGGTCGGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG	1255
QY	184	GAGGCCGTCAAGGCTCTGAAGGGCTCCAAGAAGCTGGTGCTGTCTGTACTCAGCAGGG	243
Db	1256	GAGGCCGTCAAGGCTCTGAAGGGCTCCAAGAAGCTGGTGCTGTCTGTACTCAGCAGGG	1315
QY	244	CGCATCCCTGGGGCTACGTCAACCAACCATCTACACCTGGGTGGACCCCGAGGCCCGC	303
Db	1316	CGCATCTCTGGGGCTACGTCAACCAACCATCTACACCTGGGTGGACCCCGAGGCCCGC	1375
QY	304	AGCATCTCCCCACCTCGGGCTGCCCCAGCCCCACGGTGGTGCCCTGAGCAGCAGGAG	363
Db	1376	AGCATCTCCCCACCTCGGGCTGCCCCAGCCCCACGGTGGTGCCCTGAGCAGCAGGAG	1435
QY	364	GGTGACCGGAGGAGCACCTTGACCTCTGACCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTG	423
Db	1436	GGTGACCGGAGGAGCACCTTGACCTCTGACCTGCAAGGAGGGGATGAGAAAAAGGTGAGTGGG	1495
QY	424	GTGCTGGGGACGGCCGGTCCCTGGGCGCTCACGATCC	460
Db	1496	GTGGGAAAGGAGGCCAGCCTCTCAGACACCGTATTC	1532
RESULT 9			
AC131430/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 237260)
Muzny,D.Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemeleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 237260)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237260)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23664478.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMYW
Center clone name: CH230-133E19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 227218 bases at least Q40
Consensus quality: 229137 bases at least Q30
Consensus quality: 230338 bases at least Q20
Estimated insert size: 233753; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 235931: contig of 235931 bp in length
* 235932 236031: gap of unknown length
* 236032 237260: contig of 1229 bp in length.

FEATURES

source
1. .237260
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-133E19"
1. .1635
/note="wgs contig"
230853. .231688
/note="clone boundary
clone_end:T7
site:
end_sequence:BH324527"
234235. .235931
/note="wgs end_extension
clone_end:T7"
misc_feature
misc_feature
misc_feature

ORIGIN

Query Match 10.7%; Score 180; DB 2; Length 237260;
Best Local Similarity 70.0%; Pred.No. 6.5e-31;
Matches 291; Conservative 0; Mismatches 110; Indels 15; Gaps 3;
Qy 539 AGGTTGGGACCAGATTCTAGAGTGAATGGGCGGAGCTTTCTCAACATCCTACACGACG 598
|||||
Db 220212 AGGTTGGGACCAGATTCTGAGGTGAACGGGCGGAGCTTTCTCAGCATCTCGCATGACG 220153
|||||
Qy 599 AGGCTGTGAGGCTGTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGA 658
|||||
Db 220152 AGGCAGTGAAGCTGCTCAAGTCATCCCGGCACCTCATCTGACCGTGAAGGACGTCGGGA 220093
|||||
Qy 659 GGCTGCCCATGCCCCGACCACCTGTGGACGAGACCAAGTGGATCGCCAGTCCCGGATCA 718
|||||
Db 220092 GGCTGCCCATGCCCCGACCACCTGTGGACGAGACCAAGTGGATCGCCAGTCCCGGATCG 220033
|||||
Qy 719 GGGAGACCATGGCGAATCTGGGAGGGTCTGGCAGCTCTGCTCGCTCCAATCTCCAGACCC 778
|||||
Db 220032 GGGAAAGCATCACCAACTCAGCAGGGTCTGGCCTCAGGAATGG---GGTTCCGGGATCC 219977
|||||

Matches 291; Conservative 0; Mismatches 110; Indels 15; Gaps 3;			
QY	539	AGGTTGGGACCA	GTCTAGAAAGTGAATGGGGAGCTTTCTCAACATCCTACAGACG 598
Db	178021	AGGTTGGGACCA	GTCTAGAAAGTGAATGGGGAGCTTTCTCAGCATCCTGCATGACG 178080
QY	599	AGGCTGTCAGG	CTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGACGTCGGGA 658
Db	178081	AGGCAGTGAAG	CTGCTCAAGTCATCCCGCACCTCATCTGACCGTGAAGACGTCGGGA 178140
QY	659	GGCTGCCCCAT	GCCCGCACCTGTGGACGAGACCAAGTGGATCGCCAGTTCCTCCGGATCA 718
Db	178141	GGCTGCCCCAC	GCACGTACACCTGCGGACCAAGTGGATCGCCAGTTCCTCCGGATCG 178200
QY	719	GGGAGACCAT	GCGGAATCGGCAGGGTCTGGCCACTCTGCTCGCTCCAATCTCCAGACCC 778
Db	178201	GGGAAAGCAT	CACCACTCAGCAGGGTCTGGCCCTCAGGAATGG---GGTCCGGGATCC 178256
QY	779	CAGGGCCATT	CTGAAAGCCAGTATAGCTGC--CTCCCATCCCTCCACCCGCTGGCTC 836
Db	178257	TGGGGGGCTT	CTGAGAGTCAATGTCTCTCATCTGCTCCATGGGACTCCAGTCTGG 178316
QY	837	TCCTCTCAGC	CTGCAGTCCCAACACAGGGCCCTCATTTGGCAGGACATGACCTGGGCAC 896
Db	178317	TGTTCTCAGC	CTGAGATCCAC-----TCCACTCTGTGTGGTGGCACTGGTGTGGGCC 178367
QY	897	ATCCTCTCTC	TCTTGGCCTCAGTTTCCCCATGGAAGCTGAAATACACCATCCA 952
Db	178368	CTTCCCTCTC	TCTCTAGCCCCAGTGTCTCTACAGAAAGCTGAAGTGCAGCACACA 178423

RESULT 11
AL683828/c
LOCUS
DEFINITION
Mouse DNA sequence from clone RP23-340H1 on chromosome 4 Contains the 3' end of the gene for a novel collagen triple helix repeat and fibrillar collagen C-terminal domain containing protein, the Orml1, Orm2 and Orm3 genes for orosomuroid 1, 2 and 3, an orosomuroid pseudogene, the gene for a novel protein similar to human AT-hook protein AKNA, gene 4933437N03Rik, gene 1110035G07Rik (C430046P22Rik) and two CpG islands, complete sequence.

AL683828
AL683828.8 GI:21212361
HTG; 1110035G07Rik; 4933437N03Rik; C430046P22Rik; collagen; CpG island; Orml1; Orm2; Orm3; orosomuroid.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214370)
Matthews,L.
Direct Submission
Submitted (17-OCT-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk humquerry@sanger.ac.uk
On May 25, 2002 this sequence version replaced gi:20218659.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-340H1 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseq@har.mrc.ac.uk

This sequence is the entire insert of clone RP23-340H1 The true right end of clone RP23-6H1 is at 68592 in this sequence.
FEATURES
Location/Qualifiers
1. 214370
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-340H1"
/clone_lib="RPCI-23"
33. 466
/note="match: GSS: Em:AQ970708"
224. 48609
/gene="bm340H1.1"
join(<224. 268,5141. 5194,6819. 6872,7088. 7141, 11429. 11482,14299. 14352,14778. 14822,14911. 14955, 15176. 15229,15986. 16039,16652. 16705,18803. 18856, 20304. 20357,23140. 23247,27044. 27097,27252. 27359, 28291. 28344,29180. 29233,30946. 31002,31105. 31158, 31602. 31655,31880. 31933,32539. 32592,32668. 32775, 33068. 33175,35048. 35155,37066. 37119,37537. 37599, 38049. 38102,38249. 38284,38390. 38407,38794. 38829, 39742. 39768,41448. 41513,42153. 42321,43416. 43525, 44900. 45118,46780. 48609)
/gene="bm340H1.1"
/product="bm340H1.1 (novel collagen triple helix repeat and fibrillar collagen C-terminal domain containing protein)"
/note="match: cDNAs: Em:AB058773 Em:AK003879
match: ESTs: Em:BB205277 Em:BB526911 Em:AV231828
Em:AI449266 Em:AUI20087 Em:AV595500 Em:AW652868
Em:BI557682 Em:BE683275 Em:AA296378 Em:AW436183
Em:AW658049 Em:BE756878 Em:BB621398 Em:AV600232
Em:BM545183 Em:AW446832 Em:BF773650"
/evidence=not_experimental
join(<224. 268,5141. 5194,6819. 6872,7088. 7141, 11429. 11482,14299. 14352,14778. 14822,14911. 14955, 15176. 15229,15986. 16039,16652. 16705,18803. 18856, 20304. 20357,23140. 23247,27044. 27097,27252. 27359, 28291. 28344,29180. 29233,30946. 31002,31105. 31158, 31602. 31655,31880. 31933,32539. 32592,32668. 32775, 33068. 33175,35048. 35155,37066. 37119,37537. 37599, 38049. 38102,38249. 38284,38390. 38407,38794. 38829, 39742. 39768,41448. 41513,42153. 42321,43416. 43525, 44900. 45118,46780. 46926)
/gene="bm340H1.1"
/note="match: proteins: Tr:Q96JF7 Tr:Q9JLI2 Tr:Q96HCO
Tr:Q8UJ3 Sw:O93484 Sw:Q9HAA3"
/codon_start=1
/evidence=not_experimental
/product="bm340H1.1 (novel collagen triple helix repeat and fibrillar collagen C-terminal domain containing protein)"
/protein_id="CAD62259.1"
/db_xref="GI:28172191"
/translation="GEPDGPGRPVGEQGLMGFVGEKGDGRVMGPPGA
PGKGMGHPTGPGGIGNPGEFGWPPGSRGLPMRGAKHFGPRGDPGAGEQGSX
GLKRGVGRPRPQPGQQGAGRGHSGAKGFLGIPSGSPGAKLPGEPSQGPQG
VPGPPGEMGPKPFGAVGPEGLPDGSGMKDGLPLGPPGEQLIGRGEPLGDHGP
VPGDGLKGRDPPGPDGHEGKEQGLKGEDGSPGPPGTLGTPGREGPKQGEKQGR
GAKGAKHQYLGEMGIPGEPGPTGPKGSRGLGPTGAFGRMGAQGEPLAGYNG
HKGITGLPPGPKGEGKGDQGEDGKTEGPPGPPGDRGVPVGDGRGEPDGPYPGQEG

repeat_region	/note="3.0 copies 6 mer ACAAGG 27% conserved"	
repeat_region	7797. .7807	
repeat_region	/note="11.0 copies 1 mer G 22% conserved"	
repeat_region	complement(8155. .8227)	
repeat_region	/note="L2 repeat: matches 3221. .3295 of consensus"	
repeat_region	8223. .8394	
repeat_region	/gene="bM340H1.1"	
repeat_region	/note="match: GSS: Em:AZ718678"	
repeat_region	8389. .8408	
repeat_region	/note="2.0 copies 10 mer CCACATCTGA 40% conserved"	
repeat_region	8579. .8865	
repeat_region	/note="B4A repeat: matches 5. .287 of consensus"	
repeat_region	9083. .9093	
repeat_region	/note="2.2 copies 5 mer TCCTG 22% conserved"	
repeat_region	10.3%; Score 172.6; DB 10; Length 214370;	
repeat_region	Best Local Similarity 83.4%; Pred. No. 3.3e-29;	
repeat_region	Matches 196; Conservative 0; Mismatches 39; Indels 0; Gaps 0;	
QY	194	AGGCTCTCAAGGGCTCCAAGAAAGCTGGTGGTCTGTCTGTACTCAGCAGGGGGCGATCCCTG 253
Db	186488	AGGCTCTCAAGGGCTCCAAGAAAGCTGGTGGTCTGTCTGTACTCAGCTGGGGCGATCCCG 186429
QY	254	GGGGCTACTCACCACCAACATCTACACCTGGGTGGACCCGAGGGCCGACGATCTCCC 313
Db	186428	GGGGCTATGTGACCAACCAATCTACACCTGGGTGGACCCGAGGGTCGAAGCATCTCC 186369
QY	314	CACCCTCGGGCTGCCCCAGCCCCACGGTGGTCCCTGAGGCAGCAGGAGGTGACCCGA 373
Db	186368	CTCCCTCCAGCCTGCCCCAGCCCCATGGCAGCACCCCTGAGACAGCGTGAAGATGACCGAA 186309
QY	374	GGAGCACCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGCT 428
Db	186308	GGAGTACCCTCCACCTCCTGCAAGTGGAGATGAGAAAAAGGTGAGATTGACACT 186254
RESULT 12		
AF336379/c		
LOCUS	AF336379	219809 bp DNA linear HTG 02-APR-2001
DEFINITION	Mus musculus chromosome 4 clone BAC279, *** SEQUENCING IN PROGRESS	
DEFINITION	***, 3 unordered pieces.	
ACCESSION	AF336379	
VERSION	AF336379.1	GI:13507296
KEYWORDS	HTG; HTGS PHASE1.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE	1 (bases 1 to 219809)	
JOURNAL	Rump, A., Varela, A., Mburu, P., Brown, S.D.M. and Rosenthal, A.	
REFERENCE	Mouse chromosome 4 genomic sequence	
AUTHORS	Unpublished	
TITLE	2 (bases 1 to 219809)	
JOURNAL	Rump, A.	
COMMENT	Direct Submission	
COMMENT	Submitted (17-JAN-2001) Genome Analysis, Institute of Molecular	
COMMENT	Biotechnology, Beutenbergstr. 11, Jena 07745, Germany	
COMMENT	* NOTE: This is a 'working draft' sequence. It currently	
COMMENT	* consists of 3 contigs. The true order of the pieces	
COMMENT	* is not known and their order in this sequence record is	
COMMENT	* arbitrary. Gaps between the contigs are represented as	
COMMENT	* runs of N, but the exact sizes of the gaps are unknown.	
COMMENT	* This record will be updated with the finished sequence	
COMMENT	* as soon as it is available and the accession number will	
COMMENT	* be preserved.	
FEATURES	1 199354: contig of 199354 bp in length	
source	1 199355 199454: gap of unknown length	
source	1 199455 215538: contig of 16084 bp in length	
source	1 215539 215638: gap of unknown length	
source	1 215639 219809: contig of 4171 bp in length.	
source	Location/Qualifiers	
source	1..219809	
source	/organism="Mus musculus"	

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="BAC279"

ORIGIN

Query Match 10.2%; Score 171.8; DB 2; Length 219809;
Best Local Similarity 87.4%; Pred. No. 5.1e-29;
Matches 188; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 539 AGGTTGGGACCCAGATTCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGACG 598
|||||
Db 191195 AGGTTGGAGACCCAGATTCTGGAGGTGAATGGCGGAGCTTTCTCAACATCCTGCATGATG 191136
|||||

QY 599 AGGCTGTGAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTGCGGA 658
|||||
Db 191135 AGGCAGTGAAGCTGTCAAGTCATCCCGGCACCTCATCTGACGGTGAAGGACGTGCGAA 191076
|||||

QY 659 GGCTGCCCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATGCCAGTTCCTCCGGATCA 718
|||||
Db 191075 GGCTGCCCCACGCACGTACCACTGACCGTGGACCAAGTGGATGCCAGTTCCTCCGGATCG 191016
|||||

QY 719 GGGAGACCATGGCGAAGTCTGGCAGGCTCTGGCCAC 753
|||||
Db 191015 GGGAAAGCGTCGCCCACTCAGCAGGCTCTGGTCTC 190981
|||||

RESULT 13
AC015144/c
LOCUS
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
AC015144
AC015144.1 GI:6436191
HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 16325)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211077 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..16325
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN

Query Match 6.5%; Score 108.6; DB 2; Length 16325;
Best Local Similarity 60.8%; Pred. No. 1.6e-14;
Matches 177; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 415 GTGAACCTGCTGCTGGGGACGCGCGTCCCTGGCCCTCAGATCCGTGGGGAGCTGAG 474
|||||
Db 14673 GTGGAGCTGCTGATCGAACCTGGTCAGTCTCTGGCCCTGATGATCCGTGGCGGCGTGGAG 14614
|||||

QY 475 TACGGCCTTGCGATTATACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGACGCGGG 534
|||||
Db 14613 TACGGCCTGGGATCTTTGTACCCGGCTAGATAAGGACAGCGTGGCAGATCGATCCGGA 14554
|||||

QY 535 CTCAAGGTTGGGGACCAAGATTCTAGAACTGAATGGCGGAGCTTTCTCAACATCCTACAC 594
|||||
Db 14553 CTGATGATTGGGACGAGATCCTCGAGGTCAATGGGCAATCCTTCTCGATGTGACGCAC 14494
|||||

QY 595 GACGAGGCTGTGAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTC 654
|||||
Db 14493 GACGAGGCGGTGGTCAAGTTGAAGTACCACAAACGCAATGTGCTGCTGATACGTCGTCG 14434
|||||

QY 655 GGGAGGCTGCCCCATGCCCCGACCACTGTGGACGAGACCAAGTGGATCGGCC 705
|||||
Db 14433 GGCAAGGTGCCCCACTCCTGTCACCTCCATCGAGATGGAGCCCTGGGACGCC 14383
|||||

RESULT 14
AC010699
LOCUS
DEFINITION Drosophila melanogaster clone RPC198-25L10, *** SEQUENCING IN
PROGRESS ***, 39 unordered pieces.
AC010699
AC010699.2 GI:6996798
HTG; HTGS PHASE1.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 63934)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,C., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogue,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R.,
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 63934)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 18, 2000 this sequence version replaced gi:5902990.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: DRHD
Center clone name: RPC198-25L10
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy; 13% of reads
Chemistry: Dye-terminator Big Dye; 29% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 37377 bases at least Q40
Consensus quality: 44508 bases at least Q30
Consensus quality: 49829 bases at least Q20
Estimated insert size: 47935; sum-of-contigs estimation
Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is

QY 655 GGGAGGCTGCCCCATGCGCCGACCACTGTGGACGAGACCAAGTGGATCGCC 705
Db 46841 GGCAAGGTGCCCCCACTCCTGACACCTCCATCGAGATGGAGCCCTGGGACGCC 46791

Search completed: April 22, 2004, 00:07:15
Job time : 6483.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 21, 2004, 21:36:07 ; Search time 3091 Seconds
(without alignments)
3645.807 Million cell updates/sec

Title: US-10-078-090-151
Perfect score: 1326
Sequence: 1 LRRKAHEGLGFSIRGSEH.....TWANSAGSGHSARSNLQTPG 260

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10078090/runat_20042004_132810_16588/app_query.fasta_1.455
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10078090 @CGN 1 1 5265 @runat_20042004_132810_16588 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1261.5	95.1	3566	9	AB040959 Homo sapi
2	1196.5	90.2	3178	10	AK122523 Mus muscu
3	1157.5	87.3	2763	10	AY227205 Rattus no
4	951.5	71.8	2902	9	AL110228 Homo sapi
5	706	53.2	1718	6	AX714098 Sequence
6	706	53.2	1718	9	AK056190 Homo sapi
7	438	33.0	1805	9	BC029054 Homo sapi
8	438	33.0	2037	9	AK026862 Homo sapi
9	408.5	30.8	132292	9	AL138895 Human DNA
10	408.5	30.8	171627	2	AF336382 Homo sapi
11	365.5	27.6	214370	10	AL683828 Mouse DNA
12	332	25.0	333531	2	AC127863 Rattus no
13	330.5	24.9	219809	2	AF336379 Mus muscu
14	325.5	24.5	237260	2	AC131430 Rattus no
15	321.5	24.2	2213	9	AB018687 Homo sapi
16	318.5	24.0	2162	6	AR279633 Sequence
17	318.5	24.0	2162	6	BD079792 Cancer-as
18	318.5	24.0	2162	9	AF039699 Homo sapi
19	318.5	24.0	2236	6	AR279632 Sequence
20	318.5	24.0	2236	6	BD079378 Cancer-as
21	318.5	24.0	2236	6	BD079791 Cancer-as
22	318.5	24.0	2236	9	AF039700 Homo sapi
23	318.5	24.0	2289	6	AR279636 Sequence
24	318.5	24.0	2289	6	BD079795 Cancer-as
25	318.5	24.0	2409	6	AR279637 Sequence
26	318.5	24.0	2409	6	BD079796 Cancer-as
27	315.5	23.8	2283	9	AB006955 Homo sapi
28	312.5	23.6	2063	10	AF228924 Mus muscu
29	312.5	23.6	2084	10	BC010819 Mus muscu
30	312.5	23.6	3061	10	AY103465 Mus muscu
31	312.5	23.6	3084	10	AF228925 Mus muscu
32	312	23.5	837	6	AX658639 Sequence
33	291	21.9	2184	9	BC016057 Homo sapi
34	290	21.9	16325	2	AC015144 Drosophil
35	290	21.9	63934	2	AC010699 Drosophil
36	290	21.9	110000	3	AE003535_1
37	290	21.9	173509	3	AC010031 Continuation (2 of
38	290	21.9	216041	2	AC112524 Drosophil
39	287	21.6	4306	9	AK024422 Homo sapi
40	265	20.0	179862	2	BX470066 Danio rer
41	265	20.0	191463	2	BX510367 Danio rer
42	265	20.0	219657	2	BX470198 Danio rer
43	261	19.7	191463	2	BX510367 Danio rer
44	252	19.0	3027	9	AK126776 Homo sapi
45	242	18.3	7234	5	BC063721 Xenopus l

ALIGNMENTS

RESULT 1

AB040959 3566 bp mRNA linear PRI 22-FEB-2001
LOCUS Homo sapiens mRNA for KIAA1526 protein, partial cds.
DEFINITION Homo sapiens mRNA for KIAA1526 protein, partial cds.
ACCESSION AB040959
VERSION AB040959.1 GI:7959318
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Nagase,T., Kikuno,R., Ishikawa,K., Hirose,M. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XVII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
DNA Res. 7 (2), 143-150 (2000)
JOURNAL 20277482
MEDLINE 10819331
PUBMED
REFERENCE 2 (bases 1 to 3566)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914)

FEATURES
Source

1..3566
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fj04743"
/clone_lib="pBluescriptII SK plus"
1..3566
/gene="KIAA1526"
1..2893
/gene="KIAA1526"
/note="Start codon is not identified. fj04743 cDNA clone
for KIAA1526 has a 1196-bp insertion after the position
1335 of the sequence of KIAA1526."
/codon_start=2
/product="KIAA1526 protein"
/protein_id="BAA96050.1"
/db_xref="GI:7959319"
/translation="SPRRPRLQPLPCSPNPGSSRRRLAPGDPVAVPVPARTAAPPTP
RALRARPGLGVEMNAPLDGLSVSSSTGSLGSAAGAGGGGAGLRLLSANVRQLHQAL
TALLSEAREQFTCLNAYHARRNVFDLVRTLRLDLS PVKRRLLPMLRLVLRPSDOL
LFDQYLAEGLYLPATPYQPAWGPDSAGPGEVRLVSLRRAKAHEGLGFSIRGSEH
GVGYVSLVEPGSLAEKEGLRVGDQILRVNDKSLARVTHAEAVKALKGSKKLVLVSVS
AGRIPIGVYTNHIYTWDPQGRSISPPSGLPQHGGAALRQEGDRRSTHLLQGGDEK
KVNVLVGDGRSLGLTIRGAEYGLGIYITGVDPSGEAEGSLKVGDDQILEVNGRSFLN
ILHDEAVRLKSSRHLLITVKDVGRPLPHARTTVDTETKWIASSRIRETMAANSAGFLGDL
TTEGINKPGFYKGPAGSQVTLSSLGNTQTRVLLSEQARHLLNEQEHATMAYVLYLDEYRGG
SVSVEALVMALFKLLNTHAKFSLSEVRGTISPDQLERFDHLVLRRETESMKARQPPG
PGAGDTYSMVSYSDTGSSTGSHGTSTTVSSARNTLDEETGEAVQGNINALPDVSVDD
VRSTSQGLSSFPLPRPPLAQNDLPLQPRKLRLELQPPSSMPSCSGTVFSAPQN
RSPPAGTAPTPTSSAQDLPSPIYASVSPANPSSKRPDLDAHLALVNQHPIGFPRVQ
SPPHKSPSAEATVAGGCLPPSPSGHPDQGTGNHFVMEVHRPDPSEPDVNEVRALP
QTRTASTLSQSDSGQTLSEDSGVDAGEAEASAPGRQSVSTKSRSKSLPRNERPT
DGANKPPLLEPTSTLVRVKSATLGIIEGGANTROPPLPRIVTIQGGSAHNCGQL
KVGHVILEVNGLTIRGKEHREARIIAEAFKTKDRDYIDFLVTEFNVM"

gene
CDS

1..2893
/gene="KIAA1526"
/note="Start codon is not identified. fj04743 cDNA clone
for KIAA1526 has a 1196-bp insertion after the position
1335 of the sequence of KIAA1526."
/codon_start=2
/product="KIAA1526 protein"
/protein_id="BAA96050.1"
/db_xref="GI:7959319"
/translation="SPRRPRLQPLPCSPNPGSSRRRLAPGDPVAVPVPARTAAPPTP
RALRARPGLGVEMNAPLDGLSVSSSTGSLGSAAGAGGGGAGLRLLSANVRQLHQAL
TALLSEAREQFTCLNAYHARRNVFDLVRTLRLDLS PVKRRLLPMLRLVLRPSDOL
LFDQYLAEGLYLPATPYQPAWGPDSAGPGEVRLVSLRRAKAHEGLGFSIRGSEH
GVGYVSLVEPGSLAEKEGLRVGDQILRVNDKSLARVTHAEAVKALKGSKKLVLVSVS
AGRIPIGVYTNHIYTWDPQGRSISPPSGLPQHGGAALRQEGDRRSTHLLQGGDEK
KVNVLVGDGRSLGLTIRGAEYGLGIYITGVDPSGEAEGSLKVGDDQILEVNGRSFLN
ILHDEAVRLKSSRHLLITVKDVGRPLPHARTTVDTETKWIASSRIRETMAANSAGFLGDL
TTEGINKPGFYKGPAGSQVTLSSLGNTQTRVLLSEQARHLLNEQEHATMAYVLYLDEYRGG
SVSVEALVMALFKLLNTHAKFSLSEVRGTISPDQLERFDHLVLRRETESMKARQPPG
PGAGDTYSMVSYSDTGSSTGSHGTSTTVSSARNTLDEETGEAVQGNINALPDVSVDD
VRSTSQGLSSFPLPRPPLAQNDLPLQPRKLRLELQPPSSMPSCSGTVFSAPQN
RSPPAGTAPTPTSSAQDLPSPIYASVSPANPSSKRPDLDAHLALVNQHPIGFPRVQ
SPPHKSPSAEATVAGGCLPPSPSGHPDQGTGNHFVMEVHRPDPSEPDVNEVRALP
QTRTASTLSQSDSGQTLSEDSGVDAGEAEASAPGRQSVSTKSRSKSLPRNERPT
DGANKPPLLEPTSTLVRVKSATLGIIEGGANTROPPLPRIVTIQGGSAHNCGQL
KVGHVILEVNGLTIRGKEHREARIIAEAFKTKDRDYIDFLVTEFNVM"

ORIGIN

Alignment Scores:
Pred. No.: 1.43e-72 Length: 3566
Score: 1261.50 Matches: 250
Percent Similarity: 96.17% Conservative: 1
Best Local Similarity: 95.79% Mismatches: 9
Query Match: 95.14% Indels: 1
DB: 9 Gaps: 1

US-10-078-090-151 (1-260) x AB040959 (1-3566)

1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
596 TTGGGGCGTGCCCAAGCCACAGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGAGCAC 655
21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
656 GCGTGGGCATCTACGTGTCTGTGTGAACACAGGCTCTAGCTGAGAGGAGGACTG 715
41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
716 CCGGTGGGGACACAGATTCTCGCGGTCAACGACAAATCCCTGGCCCGGTGACCCACGCG 775
61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
776 GAGGCGGTCAAGGCTCTGAAGGCTCCAAAGAGCTGGTGTGTGTGTGTACTCAGCAGG 835
81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
836 CGCATCCCTGGGGCTACGTACCAACACATCTACATCTGGTGGACCCCGCAGGCGCGC 895
101 SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlu 120
896 AGCATCTCCCAACCTCGGGGCTGCCCCAGCCCCCGTGGTGTGTGTGTGTGTGTGTGTGT 955
121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
956 GGTGACCGGAGGAGCACCTCGACCTCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTG 1015
141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
1016 GTGCTGGGGACCGCCCGTCCCTGGGCCTCACGATCCGTGGGGAGCTGAGTACGGCCTT 1075
161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
1076 GGCATTATACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGCGGCTCAAGGT 1135
181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
1136 GGGGACCAAGATTCTAGAAGTGAATGGCGGAGCTTTCTCAACATCTCTACACGAGGCT 1195
201 ValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
1196 GTCAGGTGCTTAAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTTCGGGAGGCTG 1255
221 ProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGlu 240
1256 CCCCATGCCCGCACCACTGTGGACGAGACCAAGTGATCGCAGTTCCTCCGATCAGGAG 1315
241 ThrMetAlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrPro 259
1316 ACCATGGCGAACTCGGCAGGTTTCTTGTGGCATCTCACACAGAGAAGGAATAACAAGCCA 1375
260 Gly 260
1376 GGA 1376

RESULT 2
AK122523
LOCUS AK122523
DEFINITION Mus musculus mRNA for mKIAA1526 protein.
ACCESSION AK122523
VERSION AK122523.1 GI:28972777
KEYWORDS FLI_CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1
AUTHORS Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S.,
Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA
gene: II. The complete nucleotide sequences of 400 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

of cDNA clones randomly sampled from size-fractionated libraries
DNA Res. 10, 35-48 (2003)
2 (bases 1 to 3178)
Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
Direct Submission
Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing: Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.

FEATURES
source
Location/Qualifiers
1..3178
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbh03417"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
1..3178
/gene="mKIAA1526"
1..3178
/gene="mKIAA1526"
1..3178
/gene="mKIAA1526"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence=not_experimental
/product="mKIAA1526 protein"
/protein_id="BAC65805.1"
/db_xref="GI:28972778"
/translation="ANVRQLHQALTALLSEPEREQFTHCLNAYHARNVFDLVRTLRV
LLDSPVKRLLPMLRLVIPSRLDQFYTAEGLYLPATTPYRQPAWAPDGAQGEV
RLVSLRRAXAHEGLGFSIRGGSEHGVGYVSLVEPGLSAEKEGLRVGDQILRVNDKSL
ARVTHAEAVKALKGSKLVLSVYSAGRIPGGYVTNHIYTVDPQGRSTSPSSLSLDPQH
GSTLRQREDRRSTLHLLQSGDEKKNLVLDGRSLGLTIRGAEGYGLGIYITGVDPG
SEAESGLKVGDIQILEVNGRSFLNLLHDEAVKLLKSSRHLLITVXDVGRLPHARTVD
QTKWIASRIGESVANSAGFPGDHTEGTSKPGFYKGPAGSQVTLSLGNQTRALLDD
QARHLLTEQERATMYYLAQYRGTTISVEAMVMALFELLNTHAKFSLSEVRSTISPO
DLDRFDHLVLRREIESMKARQPPGPGVGDITYSMVSYSDTGSSTGSHGTSTTVSSARNT
LDLEGTGETTQGTNALPDVSDVDVKSPEDLPGIKPPPPPLAQHDLRLGQPRKP
GREDPAPLSSAAHGIWFSAPRNRSPPTGAPTGPSSAQDSPPSIYASISHANPSS
RKPLDTHALVNOHPIGPFRVQSPPLKSPPAETPGAGACLPSPSPSPHPDVGANQ
HFVLVEVHRPDSFDPVNEVRALPQRTSTLSQLSDSGQTLSEDSGVDAGETEASTSGR
GQRTASAKNKGKQPRTERTAEGANKPGLLEPTSTLVRKSAATLGLIAIEGGANT
RQPLPRIVTIQGGSAHNCGLKVGHVILEVNGQTLRGKEHKEAARIIAEAFKTKERD
YIDFLVTFENVML"

ORIGIN

Alignment Scores:
Pred. No.: 2,01e-68 Length: 3178
Score: 1196.50 Matches: 237
Percent Similarity: 92.72% Conservative: 5
Best Local Similarity: 90.80% Mismatches: 18
Query Match: 90.23% Indels: 1
DB: 10 Gaps: 1

US-10-078-090-151 (1-260) x AK122523 (1-3178)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 319 CTGCGGCGGCCCAAGGCCACGAGGGCTTGGGCTTCAGCATCCGCGGGGCTCGGAACAC 378

QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 379 GCGGTGGGCATCTACGTGTCTCTAGTGGAGCCGGGCTCCCTTGGCAGAGAGGAGGTG 438

QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 439 CGGGTCGGGACCAGATTCTGCGCGTCAACGATAAATCTCTAGCCCGGGGTGACCCACGCG 498

QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 499 GAGGCTGTCAAGGCTCTCAAGAGCTCCAAGAGCTGGTGTCTGTATCTCAGCTGGG 558

QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 559 CGTATCCAGGGGGCTATGTGACCAACCATCTACACTGGGTGGACCCACAGGTCGA 618

QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 619 AGCACATCCCTCCCTCCAGCTGCCCCAGCCCCCATGGCAGCACCTTGAGACAGCGTGA 678

QY 121 GlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsnLeu 140
Db 679 GATGACCGAAGGAGTACCCCTCCACCTCTGCAGAGTGGAGATGAGAAAAGGTGAACCTG 738

QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
Db 739 GTGTTGGGGAGCGGCGGTCTTTGGGCTCACGATCCGAGGTGGACGAGTACGGCCCTT 798

QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
Db 799 GGCATTTACATCACTGGTGTGGACCCAGGCTCTGAAGCAGAGAGCAGCGGCTCAAGGTT 858

QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
Db 859 GGAGACCAGATTCTGGAGGTGAATGGCGGAGCTTTCTCAACATCTCCTCATGATGAGGCA 918

QY 201 ValArgLeuLeuLysSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
Db 919 GTGAAGCTGCTCAAGTCATCCCGGCACCTCATCTGACGGTGAAGGACGTCGGAAGGCTG 978

QY 221 ProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGlu 240
Db 979 CCCACGCACGTACCCCGTGGACCAAGCAAGTGGATCGCCAGTTCCTCCGGATCGGGGAA 1038

QY 241 ThrMetAlaAsnSerAlaGly--SerGlyHisSerAlaArgSerAsnLeuGlnThrPro 259
Db 1039 AGCGTCGCCAACTCAGCAGGGTTTCCAGGGGACCAACAGAGAGGAGCAAGCAAGCCA 1098

QY 260 Gly 260
Db 1099 GGA 1101

RESULT 3
AY227205 2763 bp mRNA linear ROD 04-APR-2003
LOCUS Rattus norvegicus CASK-interacting protein CIP98 (Cip98) mRNA,
DEFINITION complete cds.
ACCESSION AY227205 GI:29373052
VERSION AY227205.1
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (sites)
AUTHORS Yap,C.C., Liang,F., Yamazaki,Y., Muto,Y., Kishida,H., Hayashida,T.,
Hashikawa,T. and Yano,R.
TITLE CIP98, a novel PDZ domain protein, is expressed in the central
nervous system and interacts with calmodulin-dependent serine
kinase
JOURNAL J. Neurochem. 85 (1), 123-134 (2003)
MEDLINE 22529415
PUBMED 12641734
REFERENCE 2 (bases 1 to 2763)
AUTHORS Yap,C.C., Liang,F., Yamazaki,Y., Muto,Y., Kishida,H., Hayashida,T.,
Hashikawa,T. and Yano,R.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2003) Brain Science Institute (BSI), RIKEN, 2-1
Hirosawa, Wako, Saitama 351-0198, Japan
FEATURES Location/Qualifiers

source 1. .2763 /organism="Rattus norvegicus" /mol_type="mRNA" /strain="WST" /db_xref="taxon:10116" gene 1. .2763 /gene="Cip98" 1. .2763 /note="98 kDa" /codon_start=1 /product="CASK-interacting protein CIP98" /protein_id="AAO72534.1" /db_xref="GI:29373053" /translation="MNAQLDGLSVSSSTGSLGSAAGGGGAGLRLLSANVRQLH QALTALLSEPEREQTHCLNAYHARNVFDLVRTLRLLDSPVKRLLPMLRLVIPRS DQLLFDQYTAEGLYLPATTPYRQPAWAPDAGPGVEVRLVSLRRAKAHEGLGFSIRGG SEHGVIYVSLVEPGSLAEKEGLRVGDQILRVNDKSLARVTHAEAVKALGSKKLVLVS VYSAGRIPGGYVTNHIYTWDPQGRSTSPSSLPHGSTLRQHEDDRRSALHLLQSGDE KKNLVLGDGRSLGLTIRGAEYGLIYITGVDPGSEAESGLKVGDDQILEVNGRSFL SILHDEAVLLKSSRHLLITVKDVRPLPHARTTVDQTKWIASRIGESITNSAGFPGD LTEEGTNKPGYKPGAGQVTLSSLGNTQTRALLDDQARHLLTEQERATMMYYLDQYRG GTISVEALVMALFELLNTHAKFSLSEVRGIIISPDLDLRFHLVLRREIESMKARQPP GPGVGDYSVMYSYSDTGSSTGSHGTSTVSSARERLLWLIDMENTLDLEGTCTTQG STNALPDVSDVDDVRSPSEDLPGIKPPPPPPPLAQGHDLRGQTRKPKPVREDSAPLSAA HSGIVFSAPNRSPPPPPGIAPTPTGFPSSARDSPPSPPIYASISHANPSSRKPLDTHL ALVNOHPGIPPRVQSPPHLKSPPAEAPGACACLPSPSPSEHADAMGANQHFLVLEVH RPDSEPDVNEVRALPQTRASTLSQSDSGQTLSSEDSGVDAGETEASTSGRGRQANT KKNKGELPOTERTTEGANKPPGLLEPTSTLIRVKSAAATLGLAIEGGANTROPPLPRI VTIQRGGAHNCQQLKVGHVILEVNGQTLRGKEHREAAARVIAEAFKTKERDYIDFLVT EFNVML"

ORIGIN Alignment Scores: Pred. No.: 5.74e-66 Length: 2763 Score: 1157.50 Matches: 232 Percent Similarity: 90.80% Conservative: 5 Best Local Similarity: 88.89% Mismatches: 21 Query Match: 87.29% Indels: 3 DB: 10 Gaps: 2 US-10-078-090-151 (1-260) x AY227205 (1-2763)

Qy 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 430 CTGCGACGCGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGCGGGGGCTCGAGCAC 489

Qy 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 490 GGCCTGGGCATCTACGTGTCTCTAGTGGAGCCAGGCTCCCTGGCAGAGAGGAAGGTTA 549

Qy 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 550 CGGGTCGGGACACAGATTCTGCGCGTCAACGATAAATCTTAGCCCCGGGTGACCCACGCG 609

Qy 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 610 GAGGCTGTCAAGGCTCTCAAGGGCTCCAAGAAGTTGGTGTCTGTGTATACTCAGCTGGG 669

Qy 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 670 CGTATACCGGGGGCTATGTACCAACCAACATCTACACCTGGGTGGACCCCGAGGTCTGA 729

Qy 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 730 AGCAGATCCCTCCCTCCAGCCTT-----CCCCATGCGACGACCCCTGAGACAGCATGAA 783

Qy 121 GlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
Db 784 GATGATCGAAGGAGTGCCCTACACCTCCTGCGAGAGTGAGATGAGAAAAAGGTGAACCTG 843

Qy 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160

Db 844 GTGCTGGGGACGCGCGGTCTCTGGGCTCACAATCCGAGGTGGAGCCGAGTACGGCCTT 903

Qy 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
Db 904 GGCATTTACATCACTGCTGTGGACCCCGGCTCTGAAGCCGAAAGCAGCGGCTCAAGGT 963

Qy 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
Db 964 GGGGACCAGATTCTGGAGGTGAACGGGCGGAGCTTTCTCAGCATCTCTGATGACGAGCA 1023

Qy 201 ValArgLeuLeuLysSerArgHisLeulleLeuThrValLysAspValGlyArgLeu 220
Db 1024 GTGAAGCTGCTCAAGTCATCCCGCACCTCATCTGACCGTGAAGGACCTCGGAGGCTG 1083

Qy 221 ProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGlu 240
Db 1084 CCCACGCACGTACACCGTGGACACAGACCAAGTGGATCGCCAGTTCCCGGATCGGGAA 1143

Qy 241 ThrMetAlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrPro 259
Db 1144 AGCATCACCAACTCAGCAGGGTTTCCAGGGGACCTCACAGAAGAGGGACAAACAGCCA 1203

Qy 260 Gly 260
Db 1204 GGA 1206

RESULT 4
HSM800885
LOCUS Homo sapiens mRNA; cDNA DKFZp434N014 (from clone DKFZp434N014);
DEFINITION partial cds.
ACCESSION AL110228
VERSION AL110228.1 GI:5817166
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2902)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-1999) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp434N014) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES
Location/Qualifiers
1..2902
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp434N014"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2902
/gene="DKFZp434N014"
<1..2173
/gene="DKFZp434N014"
/note="similarity to NY-CO-38 protein"
/codon_start=2
/product="hypothetical protein"
/protein_id="CAB53685.1"
/db_xref="GI:5817167"
/db_xref="GOA:Q9UFZ3"
/db_xref="SPTREMBL:Q9UFZ3"

/translation="KLLGSRGPRLLPPECRSVACVQALKGSKKLVLSVYSAGRIPGGY
VTNHIYTWDPQGRSISPPSGLPQPHGALRQOEGDRSTLHLLOGGDEKKVNLVLGD
GRSLGTLIRGAEYGLGIYITGVDPGSEAGSLKVGQDQILEVNGRSFLNHLDEAVR
LLKSSRLHILTVKDVGRLPARTTVDETKWIASRRITRETMANSAGFLGDLTTTEGINKP
GFYKGPAGSQVTLSSLGNQTRVLLBEEQARHLNQEQHATWAYYLDEYRGGSVSEALV
MALFKLINTHAKFSLSEVRGTISPODLERFDHLVLRREIESMKARQPPGPGAGDTYS
MVSYSDTGSSTGSHGTSTTVSSARNLTLDETGEAVQGNINALPDVSVDDVRSQSGL
SSFKPLPRPPPLAQNDLPLGQPRKLGREDLQPPSPMPCSGTVFSAPQNRSPAGTA
PTPGTSSAODLPPSSPIYASVSPANPSSKRPLDAHLALVNQHPIGPFPVRVQSPPHLKSP
SAEATVAGCLLPPSPSGHPDQGTGNQHFVMEVHRPDSBPDVNEVRALPOTRFASTL
SOLSDSGTILSEDSGVDAEAEASAPGRGRQSVSTKRSKSKELPRNERPTDGANKPPG
LLEPTSTLVVKSAATLGAIEGGANTROPLRIVTIOGGSAHNCGLKVGVHILE
VNGLTLRGKEHREARIIAEAFKTKDRDYIDFLVTFEYNVL"
2824. .2829
/gene="DKFZp434N014"
2846
/gene="DKFZp434N014"
polyA_signal
polyA_site

ORIGIN

Alignment Scores:
Pred. No.: 1.27e-52 Length: 2902
Score: 951.50 Matches: 191
Percent Similarity: 89.95% Conservative: 6
Best Local Similarity: 87.21% Mismatches: 15
Query Match: 71.76% Indels: 7
DB: 9 Gaps: 2

US-10-078-090-151 (1-260) x HSM800885 (1-2902)

Qy	43	GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla	62
Db	20	GGCCACGACTACTTCCTCCTGAGTGCCGCTTCAGTGCCC-----TGT	61
Qy	63	VallysAlaLeuLysGlySerLysLeuValLeuSerValTyrSerAlaGlyArgIle	82
Db	62	GTCCAGGCTCTGAAGGCTCCAAAGAAGCTGGTGCTGTCTGTACTCAGCGGCGCATC	121
Qy	83	ProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIle	102
Db	122	CCTGGGGCTACGTCACCAACACATCTACCTGGGTGGACCCCGCAGCGCGCATC	181
Qy	103	SerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAsp	122
Db	182	TCCCCACCCTCGGGCCTGCCCCAGCCCCACGGTGGTGCCTGAGGCAGCAGGGGTGAC	241
Qy	123	ArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsnLeuValLeu	142
Db	242	CGGAGGAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTCTG	301
Qy	143	GlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIle	162
Db	302	GGGAGCGCGCGTCCCTGGCCCTCAGCATCCGTGGGGAGCTGAGTACGGCCTGGCATT	361
Qy	163	TyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAsp	182
Db	362	TACATCACTGGCGTGACCCAGGCTCTGAAGCAGAGCAGCGGGCTCAAGGTGGGGAC	421
Qy	183	GlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArg	202
Db	422	CAGATTCTAGAAGTGAATGGGCGGAGCTTCTCAACATCCTACACGACGAGGCTGCAGG	481
Qy	203	LeuLeuLysSerSerArgHisIleLeuThrVallysAspValGlyArgLeuProHis	222
Db	482	CTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGACGCTCGGGAGGCTGCCCAT	541
Qy	223	AlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMet	242
Db	542	GCCCGCACCACTGTGGACGAGACCAAGTGATCGCCAGITCCCGGATCAGGGAGACCATG	601
Qy	243	AlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrProGly	260
Db	602	CGGAACCTCGCAGGGTTTCTTGGCGATCTCACAAACAGAGGAATAACAAGCCAGGA	658

RESULT 5
AX714098
LOCUS AX714098 1718 bp DNA linear PAT 15-APR-2003
DEFINITION Sequence 782 from Patent EPI293569.
ACCESSION AX714098
VERSION AX714098.1 GI:29889026
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
Masuho,Y.
TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 782 19-MAR-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)
FEATURES
source Location/Qualifiers
1..1718
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.: 5.37e-37 Length: 1718
Score: 706.00 Matches: 140
Percent Similarity: 95.95% Conservative: 2
Best Local Similarity: 94.59% Mismatches: 6
Query Match: 53.24% Indels: 0
DB: 6 Gaps: 0
US-10-078-090-151 (1-260) x AX714098 (1-1718)

Qy	1	LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis	20
Db	1076	TTGGGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC	1135
Qy	21	GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu	40
Db	1136	GGCGTGGGCATCTACGTGTCTCTGGTGAACCAAGGCTCTCTAGCTGAGAAGGAAGGACTG	1195
Qy	41	ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla	60
Db	1196	CGGTCGGGGACCAAGATTCTCGCGGTCAACGACAAATCCCTGGCCCGGTGACCCACGCG	1255
Qy	61	GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly	80
Db	1256	GAGGCCGTCAAGGCTCTGAAGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGG	1315
Qy	81	ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg	100
Db	1316	CGCATCTCTGGGGGTACGTCAACCAACCATCTACACCTGGGTGGACCCCGCAGGCGCGC	1375
Qy	101	SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu	120
Db	1376	AGCATCTCCCCACCCTCGGGCCTGCCCCAGCCCCACCGTGTGCTGAGGCAGCAGGAG	1435
Qy	121	GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu	140
Db	1436	GGTGACCGGAGGAGCACCCCTGCACCTCTCTGCAAGGAGGGGATGAGAAAAAGGTGAGTGGG	1495
Qy	141	ValLeuGlyAspGlyArgSerLeu	148
Db	1496	GTGGGGAAGGAGGCCAGCCTCTC	1519

RESULT 6
AK056190
LOCUS AK056190 1718 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ31628 fis, clone NT2R12003344, weakly similar

ACCESSION	to PRESYNAPTIC PROTEIN SAP97.	
VERSION	AK056190	
KEYWORDS	AK056190.1 GI:16551526	
SOURCE	oligo capping; fis (full insert sequence).	
ORGANISM	Homo sapiens (human)	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.	
JOURNAL	NEDO human cDNA sequencing project	
REFERENCE	2 (bases 1 to 1718)	
AUTHORS	Isogai,T., Otsuki,T. and Sugiyama,T.	
TITLE	Direct Submission	
JOURNAL	Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)	
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.	
FEATURES	Location/Qualifiers	
source	1. 1718	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="NT2RI2003344"	
	/cell_line="NT2"	
	/cell_type="teratocarcinoma"	
	/clone_lib="NT2RI2"	
	/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction.-majorly NT2 neuron" 650..1687	
CDS	/note="unnamed protein product"	
	/codon_start=1	
	/protein_id="BAB71117.1"	
	/db_xref="GI:16551527"	
	/translation="MNAPLDGLSVSSSTGSLGSAAGAGGGGAGLRLLSANVRQLHQ ALTALLSEAREQFTHCLNAYHARNVFDLVRTLRVLDSVPKRRLLPMLRLVIPRSD QLLFDQYTAEGLYLPATTPYRQPAWGPDSDAGPGEVRLVSLRAKHEGLGFSIRGGS EHGVIYVSLVEPGSLAEKGLRVGDQILRVNDKSLARVTHAEAVKALGSKLVLVS YSAGRISGGYVTNHIYTWDPQGRSISPPSGLPQPHGGALRQOEGDRRSTLHLQGGD EKKVSGVGKGGQPLRHLRILPNPEQQSCLEAARRGWFPCGSVFPQVCTEGWCFFFAFL FDLCSVCYNTG"	
ORIGIN		
Alignment Scores:		
Pred. No.:	5.37e-37	Length: 1718
Score:	706.00	Matches: 140
Percent Similarity:	95.95%	Conservative: 2
Best Local Similarity:	94.59%	Mismatches: 6
Query Match:	53.24%	Indels: 0
DB:	9	Gaps: 0
US-10-078-090-151 (1-260) x AK056190 (1-1718)		
QY	1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20	
Db	1076 TTGCGGGCGTGCCCAAGGCCACGAGGGCTTGGGTTTCAGCATCCGTCGGGGCTCGGAGCAC 1135	

QY	21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40	
Db	1136 GGC GTGGGCATCTACGTGTCTCTGGTGAACACGAGGCTCTCTAGCTGAGAAGGAAGGACTG 1195	
QY	41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60	
Db	1196 CGG GTCGGGACGAGATTCTCGCGCTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG 1255	
QY	61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80	
Db	1256 GAGGCCGTCAAGGCTCTGAAGGGCTCCAAGAAGCTGGTGTCTGTCTGTACTCAGCAGGG 1315	
QY	81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100	
Db	1316 CGCATCTCTGGGGGTACGTACCAACCAACATCTACACCTGGTGGACCCGACGGCCGC 1375	
QY	101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120	
Db	1376 ACCATCTCCCCACCTCGGGCTGCCCCAGCCACGGTGGTGGCTGAGGCAGCAGGAG 1435	
QY	121 GlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140	
Db	1436 GGTACCCGAGGAGCACCTGCACCTCTCTGCAGGAGGGGATGAGAAAAAGGTGAGTGGG 1495	
QY	141 ValLeuGlyAspGlyArgSerLeu 148	
Db	1496 GTGGGGAAGGAGGCCAGCCTCTC 1519	
RESULT 7		
BC029054	1805 bp	linear PRI 07-OCT-2003
LOCUS	Homo sapiens hypothetical protein FLJ23209, mRNA (cDNA clone	
DEFINITION	MGC:34150 IMAGE:5179973), complete cds.	
ACCESSION	BC029054	
VERSION	BC029054.1	GI:20810506
KEYWORDS	MGC.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	1 (bases 1 to 1805)	
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.	
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
MEDLINE	22388257	
PUBMED	12477932	
REFERENCE	2 (bases 1 to 1805)	
AUTHORS	Strausberg,R.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov	

COMMENT

Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louise
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanav
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 50 Row: a Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: "matched mRNA gi: 21914924".

FEATURES

Source

```

1. .1805
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:34150 IMAGE:5179973"
/tissue_type="Brain, Lung, Testis, adult, pooled whole"
/clone_lib="NIH MGC_115"
/lab_host="DH103"
/note="Vector: pCMV-SPORT6"
. .1805
734 CGGCGCCTGGTAGTGGAGAAGTGGGTTCAACACCCCTCCGAC-----
Db
114 GlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGly 133
Qy
      :::::|||||
      ::::|||||
776 -----ACCAGCTCAGAACATGGTGTCCGGCGC---ATCGTCCACCTATACACACCTCC 826
Db
134 AspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArg 153
Qy
      |||:::
      |||||
827 GACGAC-----TTCTGCCTGGGCTTCAACATCCGT 856
Db

```

500

```

176. .1711
/codon_start=1
/product="FLJ23209 protein"
/protein_id="AAH29054.1"
/db_xref="GI:20810507"
/db_xref="LocusID:79955"
/translation="MAQFAVGFDPGLGLDSSGSLSSLRHGLGSDSGSTATRYLL
RKQRLNGLPPRGIRASSPMGVRLINSPISANDESDDIHSVRVEKSPAGRLGFSVR
GGSEHGLGIFVSKVEEGSSAERAGLCVGDKITVNGLSLESTTMGSAVKVLTSSSRLLH
MVRMRGVRPGIKFSKSKTTWVDVNRRLVVEKGSTPSDTSSEGGVRRIVHLYTTSD
DFCLGNIRGGKEFGIGIYVSKVDHGGAEENGIKVGDDQVLAANGVRDDISHSQAVE
VLKGQTHIMLTIKETCRYPAYKEMVSEYCWLDRLNGVLQOLSPASESSSVSSCASS
APYSSGSLPMDICLGQEEPGSGPGWGWRADTAMQTEPDAGGRVETWCSVRPTVIL
RDTAIRSDGPHPGRLDLSALSESPTALLLALSRRPPPIRQSQSYLTLLWEEKQQRKEK
SGSPGKEGALQRSKTLNMLNFFKGGRQGRRLARDGRREAWTLDSGSLAKTYPRLLDIEKEM
GAT"
422. .676
/note="PDZ; Region: Domain present in PSD-95, Dlg, and
ZO-1/2. Also called DHR (Dlg homologous region) or GLGF
(relatively well conserved tetrapeptide in these domains).
Some PDZs have been shown to bind C-terminal polypeptides"
/db_xref="CCDS:smarr00228"

```

misc feature

839_ .1042

/note="PDZ; Region: Domain present in PSD-95, Dlg, and ZO-1/2. Also called DHR (Dlg homologous region) or GLGF (relatively well conserved tetrapeptide in these domains). Some PDZs have been shown to bind C-terminal polypeptides" .

/db_xref="CDD:smart00228"

RESULT 8
AK026862
LOCUS
DEFINITION Homo sapiens cDNA: FLJ23209 fis, clone ADSh00512.
ACCESSION AK026862
VERSION AK026862.1 GI:10439820

AK026862 mRNA 2037 bp linear
PRI 12-SEP-2003

ORIGIN

Alignment Scores:

Pred. No.:	1.22e-19	Length:	1805
Score:	438.00	Matches:	107
Percent Similarity:	55.13%	Conservative:	38
Best Local Similarity:	40.68%	Mismatches:	90
Query Match:	33.03%	Indels:	28
DB:	9	Gaps:	6

US-10-078-090-151 (1-260) x BC029054 (1-1805)

1 LeuArgArgAlaLysAlaHisGluGly---LeuGlyPheSerIleArgGlyGlySerGlu 19

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Departent of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES Location/Qualifiers
source 1. .2037
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ADSH00512"
/tissue_type="adipose tissue"
/clone_lib="ad"
/note="cloning vector pME18SFL3"
216. .1769
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15577.1"
/db_xref="GI:10439821"
/translation="NAQGFVAVGFDPLGLDLSGSLSSLSRHLGSDSGSTATRYLL RKQRLNLPGRGIRASSPMGRVILINSPIEANSDESDIHSVRVEKSPAGRLGFSVR GGSEHGLGIFVSKVEEGSSAERAGLCVGDKITEVNGLSLETWGSVKVLTSSRLH MVVRMRGVPGIKESKEKTTWVDVNRRLVVEKGSTPSTSDSEDGVRRIVHLYTSTD DFCLGNIRGGKEFGIGIYVSKVDHGGLAENGIKVDQVLAANGVRFDDISHSQAVE VLKGQTHIMLTIKETGRYPAYKEMVSEYCWLDRLNGLVQLSPASESSSVSSCASS APYSSGSLPSDRMDICLQEEPGSRGPGWGRADTAMQTEPDAGGRVETWCSVRPTVIL RDTAIRSDGPHPRRLDSALSSEPKTALLALSRRPPPIITRSQSYLTILWEEKQORKKE KSGSPGEKALQRSKTLNMLFFKGGROGLARDGRREAWTLDSGSLAKTYPRLDIEKE MGVSPPCCPG"
ORIGIN
Alignment Scores:
Pred. No.: 1.39e-19 Length: 2037
Score: 438.00 Matches: 107
Percent Similarity: 55.13% Conservative: 38
Best Local Similarity: 40.68% Mismatches: 90
Query Match: 33.03% Indels: 28
DB: 9 Gaps: 6
US-10-078-090-151 (1-260) x AK026862 (1-2037)
QY 1 LeuArgArgAlaLysAlaHisGluGly---LeuGlyPheSerIleArgGlyGlySerGlu 19
Db 474 GTCCGGGTGGAGAGAGTCCAGCAGGAGGCTGGCTTCACGCTGCGCGGGGCTCAGAG 533
QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
Db 534 CATGGCCTGGGCATCTTCGTCAGCAAGTGGAGGAGGACGACGAGTGCAGAGCGGGCTGGC 593
QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
Db 594 CTGTGCGTGGGGACAAGATCACGGAGGTGAATGGGCTGAGCCTGGAGAGCACCACCATG 653
QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
Db 654 GGTAGCGCGCTAAAGGTGCTGACCAGCAGCGCCGCTGCACATGATGGTTCCGGCGCATG 713
QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGly 99
Db 714 GGCCGTGTCCGGGCATCAAGTCTCCCAAGGAGAGACACCGTGGGTGGATGTGGTGAAT 773
QY 100 ArgSerIle-----SerProSerGlyLeuProGlnProHisGly 113

Db 774 CGGCGCCTGGTAGTGGAGAAAGTCCGGTTCAACACCCCTCCGAC----- 815
QY 114 GlyAlaLeuArgGlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGly 133
Db 816 -----ACCAGCTCAGAAAGATGGTGTCCGGCGC--ATCGTCCACCTATACACAACTCC 866
QY 134 AspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArg 153
Db 867 GACGAC-----TTCTGCCTGGGCTTCAACATCCGT 896
QY 154 GlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAla 173
Db 897 GGGGGCAAGGAGTTTGCGCTGGGCATCTATGTGTCCAAAGTGGACCATGGTGGGTGGCC 956
QY 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
Db 957 GAGGAGAATGGCATCAAGTAGGGGACCCAGGTCTGGCAGCCACCGGTGTGAGGTTTGAC 1016
QY 194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuLeuThr 213
Db 1017 GACATCAGCCACAGCCAGGCGGTGGAGGTGCTGTAAGGGGCCAAACGCACATCATGCTGACC 1076
QY 214 ValLysAspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIle 233
Db 1077 ATCAAGGAGACCGCGCGGTATCTCTGCCTACAAGGAGATGGTTTCTGAGTACTGCTGGCTG 1136
QY 234 -----AlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSer 251
Db 1137 GACCGACTGAGCAACGCGGTGTGTGACGAGCTGTCCCGCGCTCTGAGAGCAGCTCCAGC 1196
QY 252 AlaArgSer 254
Db 1197 GTCTCTTCG 1205
RESULT 9
AL138895/c 132292 bp DNA linear PRI 27-JUN-2001
LOCUS Human DNA sequence from clone RP11-9M16 on chromosome 9, complete
DEFINITION sequence.
ACCESSION AL138895
VERSION AL138895.16 GI:14575083
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 132292)
AUTHORS Skuce,C.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Jun 28, 2001 this sequence version replaced gi:14456172.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping

Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr9>
RP11-9M16 is from the library RPCI-11.1 constructed by the group of
Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-9M16 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-9M16 is at 132292 in this
sequence. The true left end of clone RP11-402G3 is at 124695 in
this sequence. The true right end of clone RP11-821I is at 2000 in
this sequence.

FEATURES

source	Location/Qualifiers
	1..132292
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="9"
	/clone="RP11-9M16"
	/clone_lib="RPCI-11.1"
repeat_region	1..32
	/note="AluY repeat: matches 1. .32 of consensus"
repeat_region	74..301
	/note="L1MC1 repeat: matches 5718. .5955 of consensus"
repeat_region	399..696
	/note="Alusq repeat: matches 1. .298 of consensus"
repeat_region	1186..1273
	/note="MIR repeat: matches 49. .142 of consensus"
repeat_region	2271..2435
	/note="MER63A repeat: matches 46. .206 of consensus"
repeat_region	2467..2553
	/note="MIR repeat: matches 79. .165 of consensus"
repeat_region	2587..2713
	/note="L2 repeat: matches 2622. .2740 of consensus"
repeat_region	2911..3207
	/note="AluSc repeat: matches 5. .301 of consensus"
repeat_region	3214..3262
	/note="MIR repeat: matches 115. .166 of consensus"
repeat_region	4193..4262
	/note="MIR repeat: matches 60. .130 of consensus"
repeat_region	4368..5308
	/note="CpG island"
misc_feature	/evidence=not_experimental
repeat_region	5299..5606
	/note="AlusX repeat: matches 12. .312 of consensus"
repeat_region	5686..5787
	/note="MIR repeat: matches 141. .262 of consensus"
repeat_region	6714..6799
	/note="MIR repeat: matches 20. .114 of consensus"
repeat_region	6995..7437
	/note="MLT1D repeat: matches 3. .503 of consensus"
repeat_region	7440..7502
	/note="MIR repeat: matches 82. .144 of consensus"
repeat_region	8034..8179
	/note="LTR16A repeat: matches 200. .354 of consensus"
repeat_region	8160..8271
	/note="LTR16C repeat: matches 255. .368 of consensus"
repeat_region	10037..10131
	/note="MIR repeat: matches 14. .123 of consensus"
repeat_region	10604..10711
	/note="L2 repeat: matches 2405. .2521 of consensus"
repeat_region	11151..11327
	/note="MIR repeat: matches 18. .205 of consensus"
repeat_region	11579..11883
	/note="AlusX repeat: matches 1. .304 of consensus"
repeat_region	11886..12024
	/note="MIR repeat: matches 97. .243 of consensus"
repeat_region	12598..12717
	/note="MIR repeat: matches 52. .181 of consensus"
repeat_region	13839..14002
	/note="MIR repeat: matches 80. .258 of consensus"
repeat_region	14071..14205

repeat_region	/note="MIR repeat: matches 92. .226 of consensus"
repeat_region	15216..15653
	/note="MLT2CB repeat: matches 1. .492 of consensus"
repeat_region	15702..15995
	/note="L2 repeat: matches 2372. .2688 of consensus"
repeat_region	16113..16847
	/note="L2 repeat: matches 1940. .2707 of consensus"
repeat_region	16875..16922
	/note="12 copies 4 mer atga 81% conserved"
repeat_region	17482..17573
	/note="MIR repeat: matches 125. .230 of consensus"
repeat_region	18163..18379
	/note="MIR repeat: matches 23. .241 of consensus"
repeat_region	18972..19080
	/note="MIR repeat: matches 34. .144 of consensus"
repeat_region	20542..20994
	/note="Charlie4a repeat: matches 26. .506 of consensus"
repeat_region	21421..21521
	/note="L2 repeat: matches 2045. .2158 of consensus"
repeat_region	21523..21758
	/note="AluJo repeat: matches 1. .232 of consensus"
repeat_region	22863..23243
	/note="MSTA repeat: matches 1. .426 of consensus"
repeat_region	23319..23495
	/note="MIR repeat: matches 5. .192 of consensus"
repeat_region	24438..24542
	/note="MIR repeat: matches 28. .140 of consensus"
repeat_region	24993..25526
	/note="L2 repeat: matches 2099. .2709 of consensus"
repeat_region	25630..25867
	/note="MIR repeat: matches 16. .262 of consensus"
repeat_region	25889..26233
	/note="L2 repeat: matches 2363. .2674 of consensus"
repeat_region	26257..26381
	/note="MIR repeat: matches 28. .153 of consensus"
repeat_region	26919..27000
	/note="MIR repeat: matches 64. .147 of consensus"
repeat_region	27292..27435
	/note="MIR repeat: matches 102. .254 of consensus"
repeat_region	29060..29358
	/note="AlusX repeat: matches 1. .298 of consensus"
repeat_region	29523..29664
	/note="L2 repeat: matches 2593. .2734 of consensus"
repeat_region	29669..29777
	/note="MIR repeat: matches 17. .140 of consensus"
repeat_region	30134..30243
	/note="L2 repeat: matches 2167. .2289 of consensus"
repeat_region	30460..30627
	/note="MER5B repeat: matches 1. .173 of consensus"
repeat_region	30658..30770
	/note="L2 repeat: matches 2572. .2690 of consensus"
repeat_region	31123..31357
	/note="L2 repeat: matches 2514. .2731 of consensus"
repeat_region	32095..32170
	/note="19 copies 4 mer acac 88% conserved"
repeat_region	32365..32396
	/note="16 copies 2 mer tc 84% conserved"
repeat_region	32426..32485
	/note="MIR repeat: matches 96. .152 of consensus"
repeat_region	33041..33127
	/note="MIR repeat: matches 20. .115 of consensus"
repeat_region	33494..33764
	/note="AluJo repeat: matches 4. .281 of consensus"
repeat_region	34352..34417
	/note="L2 repeat: matches 2641. .2709 of consensus"
repeat_region	34554..34848
	/note="AlusX repeat: matches 1. .293 of consensus"
repeat_region	34946..35142
	/note="MIR repeat: matches 8. .207 of consensus"
repeat_region	35436..35738
	/note="AlusX repeat: matches 1. .303 of consensus"
repeat_region	35834..35965
	/note="MIR repeat: matches 2. .147 of consensus"

repeat_region 36769..36954
/note="MER53 repeat: matches 1. .188 of consensus"
repeat_region 37658..37805
/note="MIR repeat: matches 40. .190 of consensus"
repeat_region 37865..37962
/note="MIR repeat: matches 49. .140 of consensus"
repeat_region 39025..39247
/note="MIR repeat: matches 14. .252 of consensus"
repeat_region 39184..39263
/note="L2 repeat: matches 2671. .2750 of consensus"
repeat_region 39516..39569
/note="MIR repeat: matches 98. .154 of consensus"
repeat_region 40503..40600
/note="MIR repeat: matches 30. .126 of consensus"
repeat_region 40730..40829
/note="MIR repeat: matches 46. .146 of consensus"
repeat_region 40854..40893

Alignment Scores:
Pred. No.: 1.11e-15 Length: 132292
Score: 408.50 Matches: 90
Percent Similarity: 70.55% Conservativity: 13
Best Local Similarity: 61.64% Mismatches: 32
Query Match: 30.81% Indels: 11
DB: 9 Gaps: 2

US-10-078-090-151 (1-260) x AL138895 (1-132292)

QY 3 ArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyVal 22
Db 85432 AAATCGAGGCTAGAGAGGAAGGGA-----CTTCAGGGCTGG----- 85394
QY 23 GlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeuArgVal 42
Db 85393 -----CCTCTTCTTCCATGTGGGCTCCTGAGAAAACCAAGGTCTGGCCAG 85346
QY 43 GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
Db 85345 GGGAGGGTGTCTCCACACCCCGGCACCTCCCTTCTCTAACCTTTGTATCCCTCTCT 85286
QY 63 ValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIle 82
Db 85285 CCCTAGGCTCTGAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGGCCGATC 85226
QY 83 ProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIle 102
Db 85225 CCTGGGGCTACGTCAACCAACCATCTACACCTGGGTGGACCCGCGCAGCGCCGATC 85166
QY 103 SerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAsp 122
Db 85165 TCCCCACCCCTCGGGCTGCCCGCAGCCCGCAGGTGGTGGCTGAGGAGGGGTGAC 85106
QY 123 ArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeuValLeu 142
Db 85105 CGGAGGAGCACCCCTGCACCTCTGCAAGAGGGGATGAGAAAAGGTGAGTGGGTGGG 85046
QY 143 GlyAspGlyArgSerLeu 148
Db 85045 AAAGGAGGCCAGCCTCTC 85028

RESULT 10
AF336382/c 171627 bp DNA linear HTG 02-APR-2001
LOCUS Homo sapiens chromosome 9 clone R-3183P23 map 9q33, *** SEQUENCING
DEFINITION IN PROGRESS ***, 3 unordered pieces.
ACCESSION AF336382
VERSION AF336382.1 GI:13507299
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171627)

AUTHORS Rump,A., Varela,A., Mburu,P., Brown,S.D.M. and Rosenthal,A.
TITLE Mouse chromosome 4 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171627)
AUTHORS Rump,A.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-2001) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 803: contig of 803 bp in length
* 804 903: gap of unknown length
* 904 8454: contig of 7551 bp in length
* 8455 8554: gap of unknown length
* 8555 171627: contig of 163073 bp in length.
FEATURES
Location/Qualifiers
1..171627
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q33"
/clone="R-3183P23"

ORIGIN
Alignment Scores:
Pred. No.: 1.48e-15 Length: 171627
Score: 408.50 Matches: 90
Percent Similarity: 70.55% Conservativity: 13
Best Local Similarity: 61.64% Mismatches: 32
Query Match: 30.81% Indels: 11
DB: 2 Gaps: 2

US-10-078-090-151 (1-260) x AF336382 (1-171627)

QY 3 ArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyVal 22
Db 154223 AAATCGAGGCTAGAGAGGAAGGGA-----CTTCAGGGCTGG----- 154185
QY 23 GlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeuArgVal 42
Db 154184 -----CCTCTTCTTCCATGTGGGCTCCTGAGAAAACCAAGGTCTGGCCAG 154137
QY 43 GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
Db 154136 GGGAGGGTGTCTCCACACCCCGGCACCTCCCTTCTCTAACCTTTGTATCCCTCTCT 154077
QY 63 ValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIle 82
Db 154076 CCCTAGGCTCTGAAGGGCTCCAAGAGCTGGTGTGTGTGTACTCAGCAGGGCCGATC 154017
QY 83 ProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIle 102
Db 154016 CCTGGGGCTACGTCAACCAACCATCTACACCTGGGTGGACCCCGCAGCGCCGATC 153957
QY 103 SerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAsp 122
Db 153956 TCCCCACCCCTCGGGCTGCCCGCAGCCCGCAGGTGGTGGCTGAGGAGGGGTGAC 153897
QY 123 ArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeuValLeu 142
Db 153896 CGGAGGAGCACCCCTGCACCTCTGCAAGAGGGGATGAGAAAAGGTGAGTGGGTGGG 153837
QY 143 GlyAspGlyArgSerLeu 148
Db 153836 AAAGGAGGCCAGCCTCTC 153819

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 333531)
Worley, K. C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 333531)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:22855918.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZZN
Center clone name: CH230-164M17
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 204390 bases at least Q40
Consensus quality: 205853 bases at least Q30
Consensus quality: 206819 bases at least Q20
Estimated insert size: 208567; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 333531: contig of 333531 bp in length.
FEATURES
Source
1. .333531
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-164M17"
1. .1223
/note="wgs_contig"
misc_feature
213898. .215752
/note="wgs_contig"
misc_feature
332494. .333531
/note="wgs_contig"
ORIGIN
Alignment Scores:

Pred. No.: 2.73e-10 Length: 333531
Score: 332.00 Matches: 81
Percent Similarity: 57.86% Conservatives: 11
Best Local Similarity: 50.94% Mismatches: 39
Query Match: 25.04% Indels: 28
DB: 2 Gaps: 5
US-10-078-090-151 (1-260) x AC127863 (1-333531)
Qy 62 AlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArg 81
Db 141154 TCTCTTCAGGCTCTCAAGGGCTCCAAAGTGGTGTCTGTATATCTAGTGGCGT 141213
Qy 82 IleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSer 101
Db 141214 ATACCCGGGGGCTATGTCACCAACCATCTACACCTGGTGGACCCCGGTCGAAGC 141273
Qy 102 IleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGly 121
Db 141274 ACATCCCTCCCTCCAGCCTT-----CCCCATGGCAGCACCCCTGAGACAGCATGAAGAT 141327
Qy 122 AspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeuVal 141
Db 141328 GATCGAAGGAGTGCCCTACCTCCTCGAGAGTGGAGATGAGAAAAAGGTGAGA----- 141381
Qy 142 LeuGlyAsp-----GlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyr 158
Db 141382 TCTGGCCAGGTTTAAATGGGTAAAGATTGGGGGCGCTGG----- 141420
Qy 159 GlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeu 178
Db 141421 -----GGATGCTTTGGTTTCCAGAGCTCTTG 141447
Qy 179 LysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAsp 198
Db 141448 TATACCCAGGAGCTCCCTCAAGGCTAGACGGAAGGCG-----GAGCTCCGTGAC 141498
Qy 199 GluAlaValArgLeuLeuLysSerSerArgHisLeuLeuThrValLysAspVal 217
Db 141499 CAGGCTCTATTACTTCTTTTGTAGTCTCTCAAGTTTGGTATTCCTGGCTTCTCTGTG 141555

RESULT 13
AF336379/c

LOCUS AF336379 219809 bp DNA linear HTG 02-APR-2001
DEFINITION Mus musculus chromosome 4 clone BAC279, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.

ACCESSION AF336379
VERSION AF336379.1 GI:13507296

KEYWORDS HTG; HTGS PHASE1.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 219809)

AUTHORS Rump, A., Varela, A., Mburu, P., Brown, S.D.M. and Rosenthal, A.

TITLE Mouse chromosome 4 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 219809)

AUTHORS Rump, A.

TITLE Direct Submission

JOURNAL Submitted (17-JAN-2001) Genome Analysis, Institute of Molecular

Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

COMMENT * NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 199354: contig of 199354 bp in length

* 199355 199454: gap of unknown length

* 199455 215538: contig of 16084 bp in length

AUTHORS

Muzny, D., Marie, Metzker, M., Lee, Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Banderanaika, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J.,

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.tgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 235931: contig of 235931 bp in length
* 235932 236031: gap of unknown length
* 236032 237260: contig of 1229 bp in length.
Location/Qualifiers
1. .237260
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-133E19"
1. .1635
/note="wgs contig"
230853. .231688
/note="clone_boundary"
clone_end:T7
site:
end sequence:BH324527"
234235. .235931
/note="wgs_end_extension"
clone_end:T7"

Alignment Scores:
Pred. No.: 4.94e-10 Length: 237260
Score: 325.50 Matches: 67
Percent Similarity: 89.29% Conservative: 8
Best Local Similarity: 79.76% Mismatches: 8
Query Match: 24.55% Indels: 1
DB: 2 Gaps: 1

US-10-078-090-151 (1-260) x AC131430 (1-237260)
QY 167 valAspProGlySerGluAlaGluGlySerGly---LeuLysValGlyAspGlnIleLeu 185
Db 220252 CTGGATGAGGTAGCCCATGTGACAGTCTGGGTCTTGGCAGGTGGGACCGAGATTCTG 220193
QY 186 GluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLys 205
Db 220192 GAGGTGAACGGCGGAGCTTCTCAGCATCTGTCATGACGAGGAGTGAAGTGTCTCAG 220133
QY 206 SerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeuProHisAlaArgThr 225
Db 220132 TCATCCCGGCACCTCATCTGACCGTGAAGACGTGCGGAGGCTGCCACCGACGTACC 220073
QY 226 ThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMetAlaAsnSer 245
Db 220072 ACCGTGGACCAAGCAAGTGGATCGCCAGTCCCGGATCGGGAAAGCATCACTCA 220013
QY 246 AlaGlySerGly 249
Db 220012 GCAGGGTCTGGC 220001

RESULT 15
AB018687
LOCUS
DEFINITION Homo sapiens mRNA for autoimmune enteropathy-related antigen
AIE-75, complete cds.
AB018687
ACCESSION
VERSION AB018687.1 GI:5231270
KEYWORDS autoimmune enteropathy-related antigen AIE-75; AIE-75.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (sites)
Kobayashi, I., Imamura, K., Kubota, M., Ishikawa, S., Yamada, M.,
Tonoki, H., Okano, M., Storch, W.F., Moriuchi, T., Sakiyama, Y. and
Kobayashi, K.
Identification of an autoimmune enteropathy-related 75-kilodalton
antigen
Gastroenterology 117 (4), 823-830 (1999)
99431904
PUBMED 10500064
2 (bases 1 to 2213)
Kobayashi, I.
Direct Submission
Submitted (14-OCT-1998) Ichiro Kobayashi, Hokkaido University
School of Medicine, Department of Pediatrics; North-15, West-7,
Kita-ku, Sapporo, Hokkaido 060-8638, Japan
(E-mail:ichikoba@med.hokudai.ac.jp, Tel:011-716-1161,
Fax:011-736-9267)
Location/Qualifiers
1. .2213
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue type="small intestine"
90. .1748
/codon_start=1
/product="autoimmune enteropathy-related antigen AIE-75"
/protein_id="BAA81740.1"
/db_xref="GI:5231271"
/translation="MDRKVAREFRHKVDFLIENDAEDKDYLDVLRMYHOTMDVAVLVG
DLKLVINEPSRLPLFDALRPLIPLKHQVEYDQLTPRRSRKLYEVRLDRHLHPEGLGLSV
RGLEFGCGLFISHLIKGQADSVGLQVGEIVRINGYSISSCTHEEVINLIRTKTV
SIKVRHIGLIPVKSSPDPLTWQYVDQFVSESGVSGSLSPGNRENKKEKVFISLVG
SRGLGCSISSGPIQKPGIFISHVKPGSLSAEVGLEIGDQIVEVNGVDFSNLDHKEAVN
VLKSSRLTISIVAAAGRELFTDRERLAERQELQOELLQKRLAMESNKILQEQ
QEMERQRKEIAQKAAEENERYRKEQIVEEKEFKQWEEDWGSKQELLPKTITA
EVHPVPLRKPYDQGEPELEPADLDGGTEEQEQDFRKYEEGDFPYSMFTPEQIMG
KDVRLLRKKEGSLDLAEGGVDSPIGKVVSAVVEGAAERHGGIVKGEIEMAINGK
IVTDYTLAEAEALQKAWNQGDWIDLVAVCPPEYDDELTFP"
2213
polyA_site
/note="8 a nucleotides"

ORIGIN
Alignment Scores:
Pred. No.: 5.21e-12 Length: 2213
Score: 321.50 Matches: 94
Percent Similarity: 51.88% Conservative: 44
Best Local Similarity: 35.34% Mismatches: 102
Query Match: 24.25% Indels: 27
DB: 9 Gaps: 8

US-10-078-090-151 (1-260) x AB018687 (1-2213)
QY 1 LeuArgArgAlaLysAlaHis--GluGlyLeuGlyPheSerIleArgGlyGlySerGlu 19
Db 351 GTGCGTCTGGACCGTCTGCACCCCGAAGGCTCGGCCTGAGTGTGCGTGGCTGGAG 410
QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
Db 411 TTTGGTGTGGGCTCTTCATCTCCACCTCATCAAAGCGGTCAGGCAGACAGCGTCGGG 470
QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
Db 471 CTCAGGTAGGGACGAGATCGTCCGGATCAATGGATATTCCATCTCCTCTGTACCCAT 530
QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
Db 531 GAGGAGGTCAACCTCATTCGAACCAAGAAACTGTGTCTCATCAAAGTGAGACACATC 590
QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro 97
Db 591 GGCCTGATCCCCGTGAAAAGCTCTCCTGATGAGCCCCCTCACTTGGCAGTATGTGGAT--- 647

QY	98	GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg	117
Db	648	-----CAGTTTGTTCGGAATCTGGGGCGTCCGAGGCAGCCTGGGC-----	689
QY	118	GlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLys	137
Db	690	----TCCCTCGAAATCGGAAAC-----AAGGAGAAGAAG	722
QY	138	ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu	157
Db	723	GTCTTCATCAGCCTGGTAGCTCCCGAGGCCCTTGGCTGCAGCATTTCCAGCGGCCCATC	782
QY	158	TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly	177
Db	783	CAGAAGCCTGGCATCTTTATCAGCCATGTGAAACCTGGCTCCCTGTCTGTGCTGAGGTGGGA	842
QY	178	LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis	197
Db	843	TTGGAGATAGGGGACCAGATTGTTCGAAGTCAATGGCGTCGACTTCTCTAACCTGGATCAC	902
QY	198	AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr---ValLysAsp	216
Db	903	AAGGAGGCTGTAATGTGCTGAAGAGTAGCCGAGCCTGACCATCTCCATTGTAGCTGCA	962
QY	217	ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer	236
Db	963	GCTGGCCGG-----GAGCTGTTTCATGACAGACGGGAGCGGCTGGCAGAGGCG	1010
QY	237	ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu	256
Db	1011	CGGCAGCGTGAGCT-GCAGCGGCGAGGAGCTTCTCATGCAGAAGCGGCTGGCGATGGAGTC	1069
QY	257	Gln-----ThrProGly	260
Db	1070	CAACAAGATCCTCCAGGA	1087

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 21:05:04 ; Search time 4115.23 Seconds
(without alignments)
12169.150 Million cell updates/sec

Title: US-10-078-090-48
Perfect score: 1677
Sequence: 1 gagttgcggcggtgcccaaggc.....aagcgagaccaggggggaga 1677

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_est3: *
12: gb_est4: *
13: gb_est5: *
14: gb_estfun: *
15: em_estom: *
16: em_gss_hum: *
17: em_gss_inv: *
18: em_gss_pln: *
19: em_gss_vrt: *
20: em_gss_fun: *
21: em_gss_mam: *
22: em_gss_mus: *
23: em_gss_pro: *
24: em_gss_rod: *
25: em_gss_phg: *
26: em_gss_vrl: *
27: gb_gss1: *
28: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	742.4	44.3	4569	11 BC014524	BC014524 Homo sapi
2	602	35.9	602	10 BE890168	BE890168 601513147
3	531.8	31.7	905	13 BX390092	BX390092 BX390092
4	471.8	28.1	651	12 BI732824	BI732824 603353931

5	438.4	26.1	729	14	CA321445	CA321445 UI-M-FW0-
6	432	25.8	432	9	AL049078	AL049078 DKF2P434B
7	409.8	24.4	693	10	BF969269	BF969269 602269794
8	337.4	20.1	904	14	CB209310	CB209310 AGENCOURT
9	308.4	18.4	325	10	BF953476	BF953476 RC3-NN118
10	305.4	18.2	324	10	BF953475	BF953475 RC3-NN118
11	301.8	18.0	342	10	BF953545	BF953545 RC3-NN118
12	298.6	17.8	343	10	BF953479	BF953479 RC3-NN118
13	293	17.5	308	10	BF953480	BF953480 RC3-NN118
14	293	17.5	342	10	BF953472	BF953472 RC3-NN118
15	292	17.4	323	10	BF953542	BF953542 RC3-NN118
16	282.8	16.9	347	10	BF953470	BF953470 RC3-NN118
17	272	16.2	1174	13	BU166723	BU166723 AGENCOURT
18	264.8	15.8	383	10	BB872069	BB872069 BB872069
19	255.6	15.2	675	14	CA327501	CA327501 UI-M-FY0-
20	254.8	15.2	310	10	BF953522	BF953522 RC3-NN118
21	247.6	14.8	343	14	CB693399	CB693399 AMGNNUC:S
22	244.6	14.6	362	10	BF953546	BF953546 RC3-NN118
23	220	13.1	1160	14	CF661317	CF661317 CCLM09a34
24	212.6	12.7	334	10	BF953532	BF953532 RC3-NN118
25	209.6	12.5	870	14	CD760225	CD760225 AGENCOURT
26	209.4	12.5	643	14	CD310081	CD310081 StrPu691
27	206	12.3	267	14	CA949304	CA949304 it44b01.Y
28	188.8	11.3	713	13	EX880758	EX880758 BX880758
29	188.8	11.3	736	13	EX885444	EX885444 BX885444
30	188.8	11.3	761	13	EX868932	EX868932 BX868932
31	186	11.1	799	12	BG862377	BG862377 602796268
32	185.6	11.1	534	10	AW742499	AW742499 UP58a09.Y
33	184.2	11.0	531	10	BF074654	BF074654 222121 MA
34	178	10.6	586	12	BM940881	BM940881 UI-M-CG0P
35	172.8	10.3	374	14	CB692138	CB692138 AMGNNUC:T
36	172.8	10.3	396	14	CB773455	CB773455 AMGNNUC:T
37	168	10.0	641	28	AZ625547	AZ625547 IMC465B12
38	159.2	9.5	305	9	AA938876	AA938876 op74b06.s
39	130.4	7.8	342	10	BF737050	BF737050 PM1-KT004
40	123.2	7.3	1045	29	CNS04FLL	AL287778 Tetraodon
41	117.4	7.0	147	10	BF953481	BF953481 RC3-NN118
42	117.2	7.0	485	28	BZ217432	BZ217432 CH230-250
43	117.2	7.0	730	28	BH288018	BH288018 CH230-62F
44	115.6	6.9	665	28	BH290228	BH290228 CH230-114
45	112	6.7	961	29	CNS043K2	AL272891 Tetraodon

ALIGNMENTS

RESULT 1
BC014524
LOCUS BC014524 4569 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens CASK-interacting protein CIP98, mRNA (cDNA clone IMAGE:3834205), with apparent retained intron.
ACCESSION BC014524
VERSION BC014524.1 GI:17939549
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4569)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 4569)
Strausberg,R.
Direct Submission
Submitted (24-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 14 Row: f Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: retained intron.
Location/Qualifiers
1. 4569
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3834205"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH_MGC_9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

FEATURES
source
1. 4569
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3834205"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH_MGC_9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN
Query Match 44.3%; Score 742.4; DB 11; Length 4569;
Best Local Similarity 99.9%; Pred. No. 8.4e-135;
Matches 743; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
4 TTGCGGCGTCCAAAGGCCACGAGGCTTGGCTTTCAGCATCGTGGGGCTCGGAGCAC 63
1582 TTGCGGCGTCCAAAGGCCACGAGGCTTGGCTTTCAGCATCGTGGGGCTCGGAGCAC 1641
64 GCGGTGGGCATCTACGTGCTCTGCTGGAACCAAGGCTCTCTAGCTGAGAAGGAAGGACTG 123
1642 GCGGTGGGCATCTACGTGCTCTGCTGGAACCAAGGCTCTCTAGCTGAGAAGGAAGGACTG 1701
124 CCGGTGGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGTGACCCACGCG 183
1702 CCGGTGGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGTGACCCACGCG 1761
184 GAGGCCGTCAAGGCTCTGAAGGGCTCCAAGAAGCTGGTGTCTGTGTACTCAGCAGGG 243
1762 GAGGCCGTCAAGGCTCTGAAGGGCTCCAAGAAGCTGGTGTCTGTGTACTCAGCAGGG 1821
244 CGCATCCCTGGGGCTACGTACCAACCAACATCTACACCTGGGTGGACCCCGAGGGCCGC 303
1822 CGCATCCCTGGGGCTACGTACCAACCAACATCTACACCTGGGTGGACCCCGAGGGCCGC 1881
304 AGCATCTCCCAACCTCGGGGCTGCCCGGACCCACGGTGGTGGCTGCTGAGGCAGCAGGAG 363

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT

1882 AGCATCTCCCCACCTCGGGCCTGCCCCAGCCCCACGGTGGTCCCTGAGGCAGGAG 1941
364 GGTGACCGGAGGAGACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTG 423
1942 GGTGACCGGAGGAGACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTG 2001
424 GTGCTGGGGACCGCCGTCCTGGGCCTCAGATCCGTGGGGAGCTGAGTACGGCCTT 483
2002 GTGCTGGGGACCGCCGTCCTGGGCCTCAGATCCGTGGGGAGCTGAGTACGGCCTT 2061
484 GGCATTTACATCATCTGGCTGGACCCAGGCTTGAAGCAGAAAGGCGGGCTCAAGGTT 543
2062 GGCATTTACATCATCTGGCTGGACCCAGGCTTGAAGCAGAAAGGCGGGCTCAAGGTT 2121
544 GGGGACCAAGATTCTAGAAGTGAATGGCGGAGCTTCTCAACATCCTACACGAGGCT 603
2122 GGGGACCAAGATTCTAGAAGTGAATGGCGGAGCTTCTCAACATCCTACACGAGGCT 2181
604 GTCAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTGGGAGGCTG 663
2182 GTCAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTGGGAGGCTG 2241
664 CCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCCCGGATCAGGGAG 723
2242 CCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCCCGGATCAGGGAG 2301
724 ACCATGGCGAACTCGGCAGGCTT 747
2302 ACCATGGCGAACTCGGCAGGCTT 2325

RESULT 2
BE890168
LOCUS
DEFINITION
BE890168
BE890168.1 GI:10348220
EST.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 602)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9736 row: m column: 12
High quality sequence stop: 602.
Location/Qualifiers
1. 602
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3914675"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

FEATURES
source
1. 602
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3914675"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

ORIGIN
Query Match 35.9%; Score 602; DB 10; Length 602;

REFERENCE 1 (bases 1 to 651)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11919 row: m column: 07
High quality sequence stop: 651.

FEATURES
Location/Qualifiers
1..651
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5361582"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC 94"
/note="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 28.1%; Score 471.8; DB 12; Length 651;
Best Local Similarity 87.8%; Pred. No. 6.7e-82;
Matches 526; Conservative 0; Mismatches 72; Indels 1; Gaps 1;

QY 149 TCAACGACAATCCCTGGCCCGGGTGACCCACGCGAGGCGCTCAAGGCTCTGAAGGCT 208
Db 1 TCAACGATAAATCTCTAGCCCGGGTGACCCACGCGAGGCTCTCAAGGCTCTCAAGGCT 60

QY 209 CCAAGAGCTGGTGTCTGTGTACTCAGCAGGGCGCATCCCTGGGGCTACGTACCA 268
Db 61 CCAAGAGCTGGTGTCTGTATACTCAGCTGGCGTATCCC-AGGGCTATGTGACCA 119

QY 269 ACCACATCTACACCTGGGTGGACCCGAGGGCCGAGCATCTCCCCACCCCTCGGGCTGC 328
Db 120 ACCACATCTACACCTGGGTGGACCCACAGGTCGAAGCAGATCCCTCCCTCCAGCTGC 179

QY 329 CCCAGCCCCAGTGGTGCCTGTAGGAGCAGGAGGTGACCGGAGGACACCCCTGCACC 388
Db 180 CCCAGCCCCAGTGGCAGCACCTGTAGACAGCGTGAAGATGACCGAAGAGTACCTCCACC 239

QY 389 TCCTGCAAGGAGGGATGAGAAAAGTGAACTGGTGTGGGGACGGCCGTCCTGG 448
Db 240 TCCTGCAGAGTGGAGATGAGAAAAGTGAACTGGTGTGGGGACGGCCGTCCTGG 299

QY 449 GCCTCAGCATCCGTGGGGAGCTGAGTACGGCCCTTGGCATTATCATCACTGGCGTGACC 508
Db 300 GCCTCAGCATCCGAGGTGGAGCAGAGTACGGCCCTTGGCATTATCATCACTGGTGGACC 359

QY 509 CAGGCTCTGAAGCAGAGGCGAGCGGGCTCAAGGTTGGGGACCAAGTCTAGAAGTGAATG 568
Db 360 CAGGCTCTGAAGCAGAGGAGCAGCGGCCCTCAAGGTTGGAGACCAGATCTGGAGGTGAATG 419

QY 569 GCGGAGCTTTCTCAACATCTTACACGACGAGGCTGTTCAGGCTGTGTTAAGTCACTCGGC 628
Db 420 GCGGAGCTTTCTCAACATCTTCTGATGATGAGGAGTGAAGTGTCTCAAGTCACTCCCGGC 479

QY 629 ACCTCATCTGACAGTGAAGGACGTGCGGAGGCTGCCCCATGCCCCGACCACTGTGGACG 688
Db 480 ACCTCATCTGACGCTGAAGGACGTGCGGAGGCTGCCCCACGCACTACCGCTGGACC 539

QY 689 AGACCAAGTGGATCGCCAGTTCCTGGATCAGGAGACCATGGCGAACTCGGAGGGTCT 747
Db 540 AGACCAAGTGGATCGCCAGTTCCTGGATCGGGGAAAGCGTCGCCAACTCAGCAGGGTTT 598

RESULT 5
CA321445 729 bp mRNA linear EST 09-JUL-2003
LOCUS UI-M-FW0-cbz-b-04-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
DEFINITION IMAGE:6816389 5', mRNA sequence.
CA321445
ACCESSION CA321445
VERSION CA321445.1 GI:24539543
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 729)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
Location/Qualifiers
1..729
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6816389"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FW0"
/note="Organ: Brain; Vector: pYX-Asc; Site:1: EcoR I; Site:2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

The following repetitive elements were found in this cDNA

sequence: 146-183, >GC-rich#low_complexity

Seq primer: pYX-5.

FEATURES
source

ORIGIN

Query Match 26.1%; Score 438.4; DB 14; Length 729;
Best Local Similarity 89.4%; Pred. No. 2.3e-75;
Matches 472; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 5 TGCGGCGTGCCAAAGGCCACGAGGGCTTGGCTTCAGCATCCGTGGGGCTCGGAGCACG 64
Db 202 TGCGGCGCGCCAAAGGCCACGAGGGCTTGGCTTCAGCATCCGCGGGGCTCGGAACACG 261

QY 65 GCGTGGGCATCTACGTGTCTCTGTGGTGGACCGAGGCTCTCTAGTCTAGAAGGAGGACTGC 124

Db 262 GCGTGGGCATCTACGTGTCTCTAGTGGAGCCGGGCTCCCTGGCAGAGAAGGAGGTTGC 321

QY 125 GGGTCGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGTGACCCACGCGG 184

Db 322 GGGTCGGGACCAAGATTCTGCGCGTCAACGATAAATCTCTAGCCCGGTGACCCACGCGG 381

QY 185 AGGCCGCTCAAGGCTCTGAAGGGCTCCAAGAGTGGTGGTCTGTCTGTGTACTCAGCAGGGC 244

Db 382 AGGCTGTCAAGGCTCTCAAGGGCTCCAAGAGTGGTGGTCTGTCTGTATACTCAGTGGGC 441

QY 245 GCATCCCTGGGGCTACGTACCAACCAACCATCTACACCTGGGTGGACCCGAGGGCCGCA 304

Db 442 GTATCCCAGGGGGCTATGTGACCAACCAACATCTACACCTGGGTGGACCCACAGGTGAA 501

QY 305 GCATCTCCCCACCCCTGGGCGCTGCCCCAGCCCCACCGTGGTGGTGGTGGTGGTGGTGG 364

Db 502 GCATCTCCCCCTCCCTCCAGCTGCCCCAGCCCCATGGCAGCACCCCTGAGACAGCGTGAAG 561

QY 365 GTGACCGGAGGAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGG 424

Db 562 ATGACCGAAGGAGTACCCCTCACCTCCTGCAGAGTGGAGATGAGAAAAAGGTGAACCTGG 621

QY 425 TGCTGGGGGACGGCGCGTCCCTGGGCCTCACGATCCGTGGGGGAGCTGAGTACGGCCTTG 484

Db 622 TGTTGGGGACGGCCGGTCCCTGGGCCTCACGATCCGAGGTGGAGCAGAGTACGGCCTTG 681

QY 485 GCATTTACATCACTGGCGTGAGCCAGGCTCTGAAGCAGAAGGCAGCG 532

Db 682 GCATTTACATCACTGGTGTGGACCCAGGCTCTGAAGCAGACAGCAGCG 729

RESULT 6

AL049078

LOCUS

DEFINITION

DKFP434B1519_r1 434 (synonym: htes3) Homo sapiens cDNA clone

DKFP434B1519, mRNA sequence.

AL049078

AL049078.1 GI:4728387

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 432)

Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and Wiemann,S.

EST (Ottenwaelder, et al.)

Unpublished (1999)

Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No s1 sequence available.

This clone (DKFP434B1519) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source

1. .432

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFP434B1519"

/tissue_type="testis"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="434 (synonym: htes3)"

/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

ORIGIN

Query Match

Best Local Similarity 25.8%; Score 432; DB 9; Length 432;

Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 883 CATGACCTGGGCACATCCCTCTCCTCTCTTTGGCCTCAGTTTCCCCATGGAAGCTGAAAT 942

Db 1 CATGACCTGGGCACATCCCTCTCCTCTCTTTGGCCTCAGTTTCCCCATGGAAGCTGAAAT 60

QY 943 ACACCATCCAACCTGTCTCATCTCTTTATTTGTCCCCAAATTACTTAACCTCATCTATAGAC 1002

Db 61 ACACCATCCAACCTGTCTCATCTCTTTATTTGTCCCCAAATTACTTAACCTCATCTATAGAC 120

QY 1003 CTTAGTTGCTTTCATCCAAAAGTGGGACCATAAACCTGCCCTCATCCCAGATCTGTGCA 1062

Db 121 CTTAGTTGCTTTCATCCAAAAGTGGGACCATAAACCTGCCCTCATCCCAGATCTGTGCA 180

QY 1063 GATGAAAAGAGAGGGAGGGAGAGGAAAGAGAGATGCTTTGGGGTGTATTTGGCCAGAG 1122

Db 181 GATGAAAAGAGAGGGAGGGAGAGGAAAGAGAGATGCTTTGGGGTGTATTTGGCCAGAG 240

QY 1123 GCCACAGGCTGGATCCCATGAAGAAATCTGGGTGACAGGGTCTTAAAGTCATAAACTGA 1182

Db 241 GCCACAGGCTGGATCCCATGAAGAAATCTGGGTGACAGGGTCTTAAAGTCATAAACTGA 300

QY 1183 GATCCAGTTGCCAGGTGGCTGCATAGTTGCCAACAGTGAATGTGTCAACCTTTTGATCTT 1242

Db 301 GATCCAGTTGCCAGGTGGCTGCATAGTTGCCAACAGTGAATGTGTCAACCTTTTGATCTT 360

QY 1243 CATCAGAAATCTCAGGCTGGTGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTG 1302

Db 361 CATCAGAAATCTCAGGCTGGTGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTG 420

QY 1303 TCTGTCTGTCTG 1314

Db 421 TCTGTCTGTCTG 432

RESULT 7

BF969269

LOCUS

DEFINITION

BF969269

BF969269.1 GI:12336484

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 693)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM995 row: k column: 23

High quality sequence stop: 643.

Location/Qualifiers

1. .693

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4358158"

/tissue_type="adrenal cortex carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_84"

FEATURES

source

/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 24.4%; Score 409.8; DB 10; Length 693;
Best Local Similarity 95.0%; Pred. No. 9e-70;
Matches 434; Conservative 0; Mismatches 22; Indels 1; Gaps 1;
QY 4 TTGCGCGGTGCCAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 63
Db |||||||
QY 36 TTGCGCGGTGCCAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 95
Db |||||||
QY 64 GCGGTGGGCATCTACGTGTCTCTGGTGAACCAAGGCTCTCTAGCTGAGAAGGAAGGACTG 123
Db |||||||
QY 96 GCGGTGGGCATCTACGTGTCTCTGGTGAACCAAGGCTCTCTAGCTGAGAAGGAAGGACTG 155
Db |||||||
QY 124 CGGGTGGGGACCAAGTCTTGGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG 183
Db |||||||
QY 156 CGGGTGGGGACCAAGTCTTGGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG 215
Db |||||||
QY 184 GAGGCGGTCAAGGCTCTGAAGGGTCCAAAGAAGCTGGTGTCTGTGTACTCAGCAGGG 243
Db |||||||
QY 216 GAGGCGGTCAAGGCTCTGAAGGGTCCAAAGAAGCTGGTGTCTGTGTACTCAGCAGGG 275
Db |||||||
QY 244 CGCATCCCTGGGGCTACGTCACCAACCAATCTACACCTGGTGGACCCCGCAGGCCCGC 303
Db |||||||
QY 276 CGCATCCCTGGGGCTACGTCACCAACCAATCTACACCTGGTGGACCCCGCA-GGCCGC 334
Db |||||||
QY 304 AGCATCTCCCAACCTCGGGCTCGGGCTGACCCACGCTGGTGGACCCCGCAGGAG 363
Db |||||||
QY 335 AGCATCTCCCAACCTCGGGCTCGGGCTGACCCACGCTGGTGGACCCCGCAAGGAG 394
Db |||||||
QY 364 GGTGACCGGAGGAGCACCTGACCTCTGCAAGGAGGGGATGAGAAAAGGTGACCTG 423
Db |||||||
QY 395 GGTGACCGGAGGAGCACCTGACCTCTGCAAGGAGGGGATGAGAAAAGGTGAGTGGG 454
Db |||||||
QY 424 GTGCTGGGGACCGCGGTCCCTGGGCTCAGCATCC 460
Db |||||||
QY 455 GTGGGAAAGGAGGCCAGCCTCTCAGACACCGTATTTC 491
Db |||||||

RESULT 8

CB209310 904 bp mRNA linear EST 05-FEB-2003
LOCUS
DEFINITION
AGENCOURT 11348497 NIH_MGC_164 Mus musculus cDNA clone
IMAGE:30243230 5', mRNA sequence.

CB209310
CB209310.1 GI:28250873

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 904)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: NDAM0314 row: c column: 15

High quality sequence start: 18

High quality sequence stop: 487.

Location/Qualifiers

FEATURES

source

1. .904
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30243230"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_164"
/note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
Non-normalized full-length enriched library from pooled
mouse embryonic limb, maxilla and mandible, day 10.5 and
11.5 (size selected for the 0.5-1 kb fragments) Cloned
directionally, priming method: Oligo-dT. cDNA enrichment:
>1k bp, Average insert size 1.8k bp. Priming sequence:
5'GACTAGTCTAGATCGGAGCGGCGCC(T) 3'. Tissue contributed
by, David Rowe. Library constructed by ResGen, Invitrogen
Corp."

ORIGIN

Query Match 20.1%; Score 337.4; DB 14; Length 904;
Best Local Similarity 87.8%; Pred. No. 1.3e-55;
Matches 368; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 129 CGGGACCAAGATTCTGCGCTCAACGACAAATCCCTGGCCCGGGTGACCCACGCGGAGGC 188
Db |||||||
QY 21 CGGGATCAGATTCTGCGCTCAACGATAAATCTCTAGCCCGGGTGACCCACGCGGAGGC 80
Db |||||||
QY 189 CGTCAAGCTCTGAAGGGCTCAAGAAGCTGGTGTCTGTGTACTCAGCAGGGCGCAT 248
Db |||||||
QY 81 TGTCAAGCTCTCAAGGCTCAAGAAGCTGGTGTCTGTGTACTCAGCTGGCGTAT 140
Db |||||||
QY 249 CCCTGGGGCTACGTCACCAACCAATCTACACCTGGGTGGACCCCGCAGGCAT 308
Db |||||||
QY 141 CCCAGGGGCTATGTGACCAACCAATCTACACCTGGGTGGACCCCGCAGGCAT 200
Db |||||||
QY 309 CTCGCCACCTCGGGCTGCCCCAGCCCCACCGTGGTGGCTGAGGCAGGAGGGTGA 368
Db |||||||
QY 201 ATCCCTCCCTCAGCTGCCCCAGCCCCATGGCAGCACCCCTGAGACAGCTGAAGATGA 260
Db |||||||
QY 369 CCGGAGGAGCACCTGACCTCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTGGTGT 428
Db |||||||
QY 261 CCGAAGGAGTACCTCCACCTCTCTGAGAGTGGAGATGAGAAAAGGTGAACCTGGTGT 320
Db |||||||
QY 429 GGGGACCGCGCTCCCTGGGCTCAGCATCCGTGGGGAGCTAGTACGGCTTGGCAT 488
Db |||||||
QY 321 GGGGACCGCGCTCCCTGGGCTCAGCATCCGATCCGAGGTGGAGCAGAGTACGGCTTGGCAT 380
Db |||||||
QY 489 TTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGG 547
Db |||||||
QY 381 TTACATCACTGGTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTGATGG 439
Db |||||||

RESULT 9

BF953476/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BF953476 325 bp mRNA linear EST 22-JAN-2001
RC3-NN1182-121100-011-g08 NN1182 Homo sapiens cDNA, mRNA sequence.
BF953476
BF953476.1 GI:12370751
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 325)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

TITLE

JOURNAL

MEDLINE

PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC3&t2=RC3-NN1182-
121100-011-g02&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 325.
Location/Qualifiers
1. .325

FEATURES
source
1. .325
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Query Match 18.4%; Score 308.4; DB 10; Length 325;
Best Local Similarity 99.7%; Pred. No. 6e-50;
Matches 309; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 375 GAGCACCTGCACCTCCTGCAAGAGGGGATGAGAAAAGGTGAACCTGGTGTGGGGGA 434
Db |||||
QY 435 CGGCCGTCCTGGGCTCAGATCCGTGGGGAGCTGAGTACGGCCTTGGCATTACAT 494
Db |||||
QY 495 CACTGGCTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGGACCAGAT 554
Db |||||
QY 555 TCTAGAAGTGAATGGCGGAGCTTCTCAACATCCTACACGAGGCTGTAGGCTGCT 614
Db |||||
QY 615 TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGAGGCTGCCCCATGCCCG 674
Db |||||
QY 675 CACCACTGTG 684
Db 21 CACCATTTGTG 12

RESULT 10
BF953475
LOCUS RC3-NN1182-121100-011-g02 NN1182 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF953475
ACCESSION BF953475.1 GI:12370750
VERSION BF953475.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 324)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2002663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC3&t2=RC3-NN1182-
121100-011-g02&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 324.
Location/Qualifiers
1. .324

FEATURES
source
1. .324
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Query Match 18.2%; Score 305.4; DB 10; Length 324;
Best Local Similarity 98.1%; Pred. No. 2.3e-49;
Matches 309; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 365 GTGACCGGAGGAGCACCTGCACCTCCTGCAAGGAGGGGATGAGAAAAGGTGAACCTGG 424
Db |||||
QY 425 TGCTGGGGACGCGCGGTCCCTGGGCTCAGATCCGTGGGGAGCTGAGTACGGCCTTG 484
Db |||||
QY 485 GCATTTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTG 544
Db |||||
QY 545 GGGACCCAGATTCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGAGGCTG 604
Db |||||
QY 605 TCAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTGGGAGGCTGC 664
Db |||||
QY 665 CCCATGCCCCGACCA 679
Db |||||
QY 679 CCCATGCCCCGACCA 324

RESULT 11
BF953545
LOCUS RC3-NN1182-121100-012-h04 NN1182 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF953545
342 bp mRNA linear EST 22-JAN-2001

ACCESSION BF953545.1 GI:12370820
VERSION BF953545.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 342)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC3&t2=RC3-NN1182-
121100-012-h04&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 342.
Location/Qualifiers
1..342
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 18.0%; Score 301.8; DB 10; Length 342;
Best Local Similarity 97.8%; Pred. No. 1.2e-48;
Matches 306; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
365 GTGACCGGAGGAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGG 424
30 GTCACAACTTGAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGG 89
425 TGCTGGGGACGGCCGGTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGCCTTG 484
90 TGCTGGGGACGGCCGGTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGCCTTG 149
485 GCATTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTG 544
150 GCATTACATCACTGGCGTGGACCCAGGATCTGAAGCAGAGGAGCGGGCTCAAGGTTG 209
545 GGGACCAGATTCTAGAAGTGAATGGCGGAGCTTTCTCAACATCTTACACGACGAGGCTG 604
210 GGGACCAGATTCTAGAAGTGAATGGCGGAGCTTTCTCAACATCTTACACGACGAGGCTG 269
605 TCAGGCTGCTTAAGTCATCTCGGCACCTCATCTCGGACAGTGAAGGACGTCGGGAGGCTGC 664
270 TCAGGCTGCTTAAGTCATCTCGGCACCTCATCTCGGACAGTGAAGGACGTCGGGAGGCTGC 329

QY 665 CCCATGCCCGCAC 677
Db 330 CCCATGCCCGCAC 342
RESULT 12
BF953479
LOCUS BF953479 343 bp mRNA linear EST 22-JAN-2001
DEFINITION RC3-NN1182-121100-011-h06 NN1182 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF953479
VERSION BF953479.1 GI:12370754
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 343)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC3&t2=RC3-NN1182-
121100-011-h06&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 343.
Location/Qualifiers
1..343
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 17.8%; Score 298.6; DB 10; Length 343;
Best Local Similarity 98.7%; Pred. No. 4.9e-48;
Matches 301; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
375 GAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGGGGGA 434
39 GAGCACCTGACACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGGGGGA 98
435 CGGCCGGTCCCTGGGCTTCACGATCCGTGGGGGAGCTGAGTACGGCCTTGGCATTTACAT 494
99 CGGCCGGTCCCTGGGCTTCACGATCCGTGGGGGAGCTGAGTACGGCCTTGGCATTTACAT 158
495 CACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGGACCGAT 554

Db	159	CACTGGCGTGGACCCAGGCTCTGAAGCAGAAGGCAGCGGGCTCAAGGTTGGGACCAGAT	218
QY	555	TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGACGAGGCTGTCAGGCTGCT	614
Db	219	TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGACGAGGCTGTCAGGCTGCT	278
QY	615	TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGCCCCATGCCCG	674
Db	279	TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGCCCCATGCCCG	338
QY	675	CACCA 679	
Db	339	CACCA 343	
RESULT 13	BF953480/c		
LOCUS	BF953480	308 bp	mRNA linear EST 22-JAN-2001
DEFINITION	RC3-NN1182-121100-011-h10 NN1182 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	BF953480		
VERSION	BF953480.1	GI:12370755	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 308)		
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
MEDLINE	20202663		
PUBMED	10737800		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC3-NN1182-121100-011-h10&t3=2000-11-12&t4=1) Seq primer: puc 18 forward High quality sequence stop: 308. Location/Qualifiers 1. 308 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_lib="NN1182" /note="Organ: nervous normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
FEATURES	source		
ORIGIN	Query Match 17.5%; Score 293; DB 10; Length 308; Best Local Similarity 99.7%; Pred. No. 6.1e-47; Matches 304; Conservative 0; Mismatches 0; Indels 1; Gaps 1;		

QY	375	GAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGGGGGA	434
Db	304	GAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGGGGGA	245
QY	435	CGGCCGGTCCCTGGGCTTCCACGATCCGTGGGGGAGCTGAGTACGGCTTGGCATTTACAT	494
Db	244	CGGCCGGTCCCTGGGCTTCCACGATCCGTGGGGGAGCTGAGTACGGCTTGGCATTTACAT	185
QY	495	CACTGGCGTGGACCCAGGCTCTGAAGCAGAAGGCAGCGGGCTCAAGTTGGGGACCAGAT	554
Db	184	CACTGGCGTGGACCCAGGCTCTGAAGCAGAAGGCAGCGGGCTCAAGTTGGGGACCAGAT	125
QY	555	TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGACGAGGCTGTCAGGCTGCT	614
Db	124	TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGACGAGGCTGTCAGGCTGCT	65
QY	615	TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGTCGCCCATGCCCG	674
Db	64	TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGAC-TCGGGAGGTCGCCCATGCCCG	6
QY	675	CACCA 679	
Db	5	CACCA 1	
RESULT 14	BF953472		
LOCUS	BF953472	342 bp	mRNA linear EST 22-JAN-2001
DEFINITION	RC3-NN1182-121100-011-d05 NN1182 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	BF953472		
VERSION	BF953472.1	GI:12370747	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 342)		
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
MEDLINE	20202663		
PUBMED	10737800		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC3-NN1182-121100-011-d05&t3=2000-11-12&t4=1) Seq primer: puc 18 forward High quality sequence start: 8 High quality sequence stop: 342. Location/Qualifiers 1. 342 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_lib="NN1182" /note="Organ: nervous normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application		
FEATURES	source		

No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 17.5%; Score 293; DB 10; Length 342;
Best Local Similarity 99.7%; Pred. No. 6.1e-47;
Matches 304; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 375 GAGCACCCCTGCACCTCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGCTGGGGGA 434
Db 39 GAGCACCCCTGCA-CTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGCTGGGGGA 97
QY 435 CGGCCGGTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGCCTTGGCATTACAT 494
Db 98 CGGCCGGTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGCCTTGGCATTACAT 157
QY 495 CACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGGACCAGAT 554
Db 158 CACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGGACCAGAT 217
QY 555 TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGAGGCTGTCAGGCTGCT 614
Db 218 TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGAGGCTGTCAGGCTGCT 277
QY 615 TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGCCCATGCCCG 674
Db 278 TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGCCCATGCCCG 337
QY 675 CACCA 679
Db 338 CACCA 342

RESULT 15
BF953542
LOCUS BF953542 323 bp mRNA linear EST 22-JAN-2001
DEFINITION RC3-NN1182-121100-012-f04 NN1182 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF953542
VERSION BF953542.1 GI:12370817
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 323)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC3&t2=RC3-NN1182-
121100-012-f04&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 323.
Location/Qualifiers

FEATURES

source
1. .323
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 17.4%; Score 292; DB 10; Length 323;
Best Local Similarity 100.0%; Pred. No. 9.6e-47;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 388 CTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGCTGGGGACGGCCGCTCCCTG 447
Db 32 CTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCTGGTGCTGGGGACGGCCGCTCCCTG 91
QY 448 GGCCTCACGATCCGTGGGGAGCTGAGTACGGCCTTGGCATTACATCAGCTGGCGTGGAC 507
Db 92 GGCCTCACGATCCGTGGGGAGCTGAGTACGGCCTTGGCATTACATCAGCTGGCGTGGAC 151
QY 508 CCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGACCAGATTCTAGAAGTGAAT 567
Db 152 CCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGACCAGATTCTAGAAGTGAAT 211
QY 568 GGGCGGAGCTTTCTCAACATCCTACACGAGGCTGTCAGGCTGCTTAAGTCATCTCGG 627
Db 212 GGGCGGAGCTTTCTCAACATCCTACACGAGGCTGTCAGGCTGCTTAAGTCATCTCGG 271
QY 628 CACCTCATCCTGACAGTGAAGGACGTCGGGAGGCTGCCCCCATGCCCGCACCA 679
Db 272 CACCTCATCCTGACAGTGAAGGACGTCGGGAGGCTGCCCCCATGCCCGCACCA 323

Search completed: April 22, 2004, 01:43:32
Job time : 4129.23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 19:26:07 ; Search time 259.231 Seconds
(without alignments)
10930.600 Million cell updates/sec

Title: US-10-078-090-47
Perfect score: 667
Sequence: 1 gcgtggtcgcggcccgaggt.....aagcgagaccaggggggaga 667

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	667	100.0	667	6	ABV83604 Human bre
2	647.2	97.0	1677	6	ABV83605 Human bre
3	76.8	11.5	710	6	ABT09598 Phase-1 R
4	76.8	11.5	3316	4	ABL10742 Drosophil
5	76.6	11.5	1817	5	ABA15256 Human ner
6	75.6	11.3	271990	9	ADD25213 Fertility
7	73.6	11.0	2278	4	ABL12518 Drosophil
8	73.4	11.0	667	9	ADD19798 Seabass p
9	73.4	11.0	67832	8	ADA02801 Mouse Itk
10	73.4	11.0	67832	9	ADB72539 Mouse Itk
11	73.4	11.0	67832	9	ADC85281 Mouse Itk
12	73.2	11.0	742	6	ABS76748 Frog embr
13	73.2	11.0	42998	7	ADA14747 Human rib
14	73.2	11.0	42999	6	ABS65032 Invertebr
15	73.2	11.0	42999	9	AAD61411 Human rib
16	72.6	10.9	338	9	ADD19844 Oreochrom
17	72.4	10.9	596	9	ADD19952 Oreochrom
18	72.4	10.9	814	9	ADD20177 Oreochrom
19	72.4	10.9	41907	8	ADA37416 Human rDN
20	72.2	10.8	534	9	ADD20617 Oreochrom
21	72.2	10.8	110000	7	ABX16390_5 Continuation (6 of
22	72	10.8	448	2	AAT68682 Penaeus v
23	72	10.8	605	9	ADD20225 Oreochrom

C	24	71.8	10.8	299	9	ADD19778	Gadus mor
	25	71.8	10.8	475	4	AAI12688	Probe #26
	26	71.8	10.8	475	4	ABA54378	Human foe
	27	71.8	10.8	475	4	AAI34039	Probe #27
	28	71.8	10.8	475	4	ABA43925	Human bre
	29	71.8	10.8	475	4	ABA24157	Probe #26
	30	71.8	10.8	475	4	AAK28111	Human bon
	31	71.8	10.8	475	4	AAK02673	Human bra
	32	71.8	10.8	475	4	ABS27717	Human liv
	33	71.8	10.8	475	5	AAI02597	Probe #25
	34	71.8	10.8	475	6	ABS02604	Human gen
C	35	71.8	10.8	527	6	ABQ92801	Triticum
	36	71.8	10.8	546	4	AAI19220	Probe #91
	37	71.8	10.8	546	4	ABA64228	Human foe
	38	71.8	10.8	546	4	AAI44380	Probe #13
	39	71.8	10.8	546	4	ABA31373	Probe #98
	40	71.8	10.8	546	4	AAK38426	Human bon
	41	71.8	10.8	546	4	AAK12706	Human bra
	42	71.8	10.8	546	4	ABS38019	Human liv
	43	71.8	10.8	546	6	ABS12484	Human gen
	44	71.6	10.7	316	4	AAH29391	Drosophil
	45	71.6	10.7	48995	8	ADA02579	Mouse Rel

ALIGNMENTS

RESULT 1
ABV83604
ID ABV83604 standard; cDNA; 667 BP.
XX
AC ABV83604;
XX
DT 06-DEC-2002 (first entry)
XX
DE Human breast specific gene SEQ ID NO 47.
XX
KW Human; breast; cytostatic; cancer; transgenic; gene therapy; vaccine;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO200266605-A2.
PD 29-AUG-2002.
XX
PF 14-FEB-2002; 2002WO-US004284.
XX
PR 15-FEB-2001; 2001US-0268999P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
PI Sun Y, Liu C;
XX
DR WPI; 2002-713345/77.
XX
PT New isolated breast specific nucleic acid molecules and polypeptides,
PT useful for identifying, diagnosing, monitoring, staging, imaging and
PT treating breast cancer and non-cancerous disease states in breast tissue.
XX
PS Claim 1; Page 170-171; 254pp; English.
XX
CC The invention relates to human breast specific nucleic acids (I)
CC comprising: (a) a sequence encoding any one of 95 protein sequences
CC (ABP66614-ABP66708); (b) any one of 115 polynucleotide sequences
CC (ABV83558-ABV83672); (c) a molecule that selectively hybridizes to (a) or
CC (b); (d) a molecule having at least 60% sequence identity to (a) or (b).
CC The breast specific nucleic acid molecules, polypeptides and antibodies
CC are useful for identifying, diagnosing, monitoring, staging, imaging and
CC treating breast cancer and non-cancerous disease states in breast tissue.
CC They are also useful for producing transgenic animals and cells and
CC producing engineered breast tissue for treatment and research. The

RESULT 5
ABA15256
ID ABA15256 standard; DNA; 1817 BP.
XX
AC ABA15256;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 7587.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 21:08:41 ; Search time 280.003 Seconds
(without alignments)
10740.063 Million cell updates/sec

Title: US-10-078-090-47
Perfect score: 667
Sequence: 1 gcgtggtcgccgcccagggt.....aagcgagaccaggggggaga 667

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	667	100.0	667	15	US-10-078-090-47
2	647.2	97.0	1677	15	US-10-078-090-48
3	76.6	11.5	553	13	US-10-027-632-285026
4	76.6	11.5	553	13	US-10-027-632-285028
5	76.6	11.5	553	16	US-10-027-632-285026
6	76.6	11.5	553	16	US-10-027-632-285028
7	76.2	11.4	553	13	US-10-027-632-285027
8	76.2	11.4	553	16	US-10-027-632-285027
9	76.2	11.4	113633	13	US-10-087-192-19
C 10	75.6	11.3	271990	15	US-10-195-144-87
C 11	75.6	11.3	271990	16	US-10-345-072-87
C 12	74.6	11.2	53954	13	US-10-087-192-262
C 13	74.4	11.2	350764	13	US-10-087-192-1864
14	73.4	11.0	67832	12	US-09-997-722-67

C 15	73.4	11.0	130427	15	US-10-175-523-87	Sequence 87, Appl
C 16	73.2	11.0	742	9	US-09-910-943-3	Sequence 3, Appli
17	73.2	11.0	42999	9	US-09-799-462A-17	Sequence 17, Appl
18	73.2	11.0	42999	10	US-09-836-911A-17	Sequence 17, Appl
19	73.2	11.0	42999	10	US-09-738-630-73	Sequence 73, Appl
20	73.2	11.0	42999	14	US-10-125-767-17	Sequence 17, Appl
21	73.2	11.0	42999	15	US-10-151-081-17	Sequence 17, Appl
22	73.2	11.0	42999	15	US-10-287-313-17	Sequence 17, Appl
23	73.2	11.0	42999	15	US-10-219-694-17	Sequence 17, Appl
24	72.8	10.9	595	13	US-10-027-632-90339	Sequence 90339, A
25	72.8	10.9	595	13	US-10-027-632-317388	Sequence 317388, A
26	72.8	10.9	595	16	US-10-027-632-90339	Sequence 90339, A
27	72.8	10.9	595	16	US-10-027-632-317388	Sequence 317388,
28	72.8	10.9	83836	13	US-10-087-192-1849	Sequence 1849, Ap
29	72.6	10.9	38269	16	US-10-085-117-325	Sequence 325, App
C 30	72.6	10.9	116422	13	US-10-087-192-1195	Sequence 1195, Ap
31	72.4	10.9	126413	13	US-10-087-192-1831	Sequence 1831, Ap
C 32	72.2	10.8	117382	13	US-10-087-192-1435	Sequence 1435, Ap
33	72.2	10.8	155579	16	US-10-085-117-283	Sequence 283, App
34	72.2	10.8	263744	15	US-10-229-834A-6	Sequence 6, Appli
C 35	72.2	10.8	659158	9	US-09-771-208-20	Sequence 20, Appl
36	71.8	10.8	475	9	US-09-864-761-2623	Sequence 2623, Ap
37	71.8	10.8	546	9	US-09-864-761-9839	Sequence 9839, Ap
C 38	71.8	10.8	194049	13	US-10-087-192-1495	Sequence 1495, Ap
39	71.6	10.7	48995	12	US-10-052-482-85	Sequence 85, Appl
40	71.6	10.7	63502	13	US-10-087-192-271	Sequence 271, App
41	71.2	10.7	249487	14	US-10-026-188-3	Sequence 3, Appli
C 42	71	10.6	9450	15	US-10-126-912-2	Sequence 2, Appli
C 43	71	10.6	9757	15	US-10-126-912-7	Sequence 7, Appli
C 44	71	10.6	9828	15	US-10-126-912-1	Sequence 1, Appli
C 45	71	10.6	39924	13	US-10-087-192-1993	Sequence 1993, Ap

ALIGNMENTS

RESULT 1

US-10-078-090-47
; Sequence 47, Application US/10078090
; Publication No. US2003004815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 667
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-078-090-47

Query Match 100.0%; Score 667; DB 15; Length 667;
Best Local Similarity 100.0%; Pred. No. 6.8e-177;
Matches 667; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTGGTGGCGCCGAGGTCCATAACCCCTGCCTCATCCAGATCTGTGCAGATGAAAG 60

Db 1 GCGTGGTGGCGCCGAGGTCCATAACCCCTGCCTCATCCAGATCTGTGCAGATGAAAG 60

QY 61 AGACGGAGGGAGAGGAAAGAGAGAGATGCTTTGGGGTGTATTGGCCAGAGGCCACCAG 120

Thu Apr 22 09:46:01 2004

```

; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Priithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 87
; LENGTH: 130427
; TYPE: DNA
; ORGANISM: Mus musculus
; US-10-175-523-87

```

[illegible]

Search completed: April 22, 2004, 01:59:50
Job time : 287.003 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 21:05:04 ; Search time 1636.77 Seconds
(without alignments)
12169.150 Million cell updates/sec

Title: US-10-078-090-47
Perfect score: 667
Sequence: 1 gcgtggtcgcccgaggt.....aagcgagaccaggggggaga 667

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rod:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	284.2	42.6	432	9 AL049078	AL049078 DKFZp434B
C 2	88.6	13.3	1246	9 AL514829	AL514829 AL514829
C 3	87.8	13.2	627	28 AZ997081	AZ997081 2M0283C01
C 4	87	13.0	752	28 BZ271973	BZ271973 CH230-275

C 5	86.6	13.0	715	9 AL663389	AL663389 AL663389
C 6	84.6	12.7	492	28 AZ897298	AZ897298 RPCI-24-1
C 7	84	12.6	628	29 BX151392	BX151392 Danio rer
C 8	83.8	12.6	851	14 CA581263	CA581263 EST000938
C 9	82.4	12.4	737	29 BX146272	BX146272 Danio rer
C 10	82	12.3	796	28 BH350473	BH350473 CH230-120
C 11	81.8	12.3	583	29 CNS044QG	AL274417 Tetraodon
C 12	81.2	12.2	740	29 CNS02YRZ	AL220040 Tetraodon
C 13	80.6	12.1	564	28 BH383538	BH383538 AG-ND-119
C 14	80.6	12.1	635	28 BH370475	BH370475 AG-ND-180
C 15	80.4	12.1	602	29 CC775881	CC775881 CH240_130
C 16	79.8	12.0	595	29 DR44N19T	AL973430 Danio rer
C 17	79.4	11.9	989	29 CNS01G9Q	AL142863 Anopheles
C 18	79.2	11.9	474	12 BM582064	BM582064 170006872
C 19	79.2	11.9	801	29 CNS01VC9	AL168930 Tetraodon
C 20	79	11.8	512	29 BX242790	BX242790 Danio rer
C 21	79	11.8	575	29 BX182349	BX182349 Danio rer
C 22	78.8	11.8	704	28 AQ780204	AQ780204 HS_3138_B
C 23	78.6	11.8	594	29 DR9L1S	AL733910 Danio rer
C 24	78.4	11.8	528	13 BQ480630	BQ480630 faa92a03
C 25	78.4	11.8	543	28 BZ179552	BZ179552 CH230-444
C 26	78.4	11.8	758	28 BZ233672	BZ233672 CH230-281
C 27	78.4	11.8	880	29 BX240451	BX240451 Danio rer
C 28	78.4	11.8	1159	9 AL514869	AL514869 AL514869
C 29	78.2	11.7	534	9 AL514819	AL514819 AL514819
C 30	78.2	11.7	608	28 AZ973484	AZ973484 2M0247B14
C 31	78.2	11.7	687	28 AZ335248	AZ335248 1M0065D11
C 32	78	11.7	551	9 AI586009	AI586009 vs40e12.x
C 33	78	11.7	581	29 DR36D12T	AL986788 Danio rer
C 34	78	11.7	776	28 BZ105168	BZ105168 CH230-150
C 35	78	11.7	1353	13 BX404043	BX404043 BX404043
C 36	77.6	11.6	345	14 CB937318	CB937318 IpCGJX13
C 37	77.6	11.6	522	14 CF777662	CF777662 tad14b01
C 38	77.6	11.6	544	28 AZ736516	AZ736516 RPCI-24-7
C 39	77.4	11.6	442	29 CNS02E4P	AL193282 Tetraodon
C 40	77.4	11.6	491	28 AZ290346	AZ290346 RPCI-23-1
C 41	77.4	11.6	674	29 BX196307	BX196307 Danio rer
C 42	77.4	11.6	805	28 CC073103	CC073103 CSU-K33r
C 43	77.4	11.6	809	28 CC080479	CC080479 CSU-K33r
C 44	77.2	11.6	552	29 CNS03W90	AL263421 Tetraodon
C 45	77.2	11.6	638	28 AQ628460	AQ628460 CITBI-E1

ALIGNMENTS

RESULT 1
AL049078
LOCUS
DEFINITION DKFZp434B1519_r1 434 (synonym: htes3) Homo sapiens cDNA clone
432 bp mRNA linear EST 04-SEP-2003
DKFZp434B1519 mRNA sequence.
ACCESSION AL049078
VERSION AL049078.1 GI:4728387
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 432)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Ottenwaelder, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS

MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available.
This clone (DKFZp434B1519) is available at the RZPD in Berlin.

[illegible]

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 21, 2004, 21:20:47 ; Search time 367 Seconds
(without alignments)
3009.623 Million cell updates/sec

Title: US-10-078-090-151
Perfect score: 1326
Sequence: 1 LRRKAHEGLGFSIRGGSEH.....TMANSAGSHSARSNLQTPG 260

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues 6747726
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO spool_p/US10078090/runat 20042004 132809 16579/app query.fasta_1.455
-DB=N Geneseq_29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10078090 @CGN 1 1 708 @runat 20042004 132809 16579 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_29Jan04:
1: Geneseqn1980s:
2: Geneseqn1990s:
3: Geneseqn2000s:
4: Geneseqn2001as:
5: Geneseqn2001bs:
6: Geneseqn2002s:
7: Geneseqn2003as:
8: Geneseqn2003bs:
9: Geneseqn2003cs:
10: Geneseqn2004s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	1326	100.0	1677	6	ABV83605 Human bre
2	954	71.9	3102	3	AAC75441 Human ORF
3	706	53.2	1718	7	ADA53214 Human cod
4	438	33.0	2822	7	ABX71178 Novel hum
5	324	24.4	1338	4	ABL29731 Drosophil
6	321.5	24.2	2259	7	ABX34636 Human mdd
7	318.5	24.0	2162	2	AAX40058 Colon can
8	318.5	24.0	2236	2	AAX40057 Colon can

9	318.5	24.0	2236	2	AAX39644	Aax39644 Renal can
10	318.5	24.0	2289	2	AAX40061	Aax40061 Colon can
11	318.5	24.0	2409	2	AAX40062	Aax40062 Colon can
12	312	23.5	837	7	ABZ75586	Abz75586 SID polyp
13	290	21.9	6497	4	ABL29730	AbL29730 Drosophil
14	280	21.1	1489	4	ABL10215	AbL10215 Drosophil
15	231	17.4	3046	8	ACC85071	Acc85071 Human MBC
16	231	17.4	4108	8	AAD57342	Aad57342 Human kin
17	229	17.3	3071	7	ACA56540	Aca56540 Human sig
18	229	17.3	3071	8	ACC85072	Acc85072 Human MBC
19	228.5	17.2	3100	2	AAV20419	Aav20419 Human dis
20	228.5	17.2	3100	2	AAX86725	Aax86725 DNA seque
21	228.5	17.2	3100	8	ACC85073	Acc85073 Human MBC
22	228.5	17.2	3196	2	AAV20420	Aav20420 Human dis
23	228.5	17.2	5147	5	ABV25006	Abv25006 Human pro
24	226.5	17.1	4801	9	ADD78275	Add78275 Human CGD
25	226.5	17.1	4884	8	AAD57243	Aad57243 Human CGD
26	226.5	17.1	4933	7	ACD13338	Acd13338 Human DNA
27	226.5	17.1	5079	8	AAD57244	Aad57244 Human CGD
28	226.5	17.1	5139	7	ABV72514	Abv72514 Nucleotid
29	222	16.7	3995	6	ABK35489	Abk35489 Human end
30	222	16.7	3995	6	ABK35548	Abk35548 Gene DLG4
31	222	16.7	3995	6	ABV78023	Abv78023 Hypoxia-r
32	222	16.7	3995	8	ACC85074	Acc85074 Human MBC
33	217.5	16.4	2963	2	AAT93775	Aat93775 PSD-93 co
34	217	16.4	3671	4	ABL10214	Abl10214 Drosophil
35	214	16.1	1274	4	AAH73917	Aah73917 Human PDZ
36	210	15.8	993	3	AAA30453	Aaa30453 Human CDN
37	209	15.8	1492	9	ADB58401	Adb58401 Toxicity-
38	209	15.8	1492	9	ADB52979	Adb52979 Primary r
39	203.5	15.3	3454	4	ABL03235	AbL03235 Drosophil
40	203	15.3	4235	4	AAH99661	Aah99661 Human pro
41	203	15.3	4235	4	ABA08948	AbA08948 Human sem
42	199.5	15.0	4607	6	ABK46540	Abk46540 DNA encod
43	198	14.9	8640	6	ABS63443	Abs63443 Human CDN
44	198	14.9	8640	6	ABS63444	Abs63444 Human CDN
45	197.5	14.9	873	6	ABK46532	Abk46532 Human PDZ

ALIGNMENTS

RESULT 1
ABV83605
ID ABV83605 standard; cDNA; 1677 BP.
XX
AC ABV83605;
XX
DT 06-DEC-2002 (first entry)
XX
DE Human breast specific gene SEQ ID NO 48.
XX
KW Human; breast; cytostatic; cancer; transgenic; gene therapy; vaccine;
XX gene; ss.
XX Homo sapiens.
OS
XX WO200266605-A2.
PN
XX
PD 29-AUG-2002.
XX
PF 14-FEB-2002; 2002WO-US004284.
XX
PR 15-FEB-2001; 2001US-0268999P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Salceda S, Macina RA, Hu P, Recipon H, Karra K, Caferkey R;
PI Sun Y, Liu C;
XX
DR WPI; 2002-713345/77.
XX
PT New isolated breast specific nucleic acid molecules and polypeptides,
PT useful for identifying, diagnosing, monitoring, staging, imaging and

PT treating breast cancer and non-cancerous disease states in breast tissue.
XX
PS Claim 1; Page 171-172; 254pp; English.
XX
CC The invention relates to human breast specific nucleic acids (I)
CC comprising: (a) a sequence encoding any one of 95 protein sequences
CC (ABP6614-ABP66708); (b) any one of 115 polynucleotide sequences
CC (ABV83558-ABV83672); (c) a molecule that selectively hybridizes to (a) or
CC (b); (d) a molecule having at least 60% sequence identity to (a) or (b).
CC The breast specific nucleic acid molecules, polypeptides and antibodies
CC are useful for identifying, diagnosing, monitoring, staging, imaging and
CC treating breast cancer and non-cancerous disease states in breast tissue.
CC They are also useful for producing transgenic animals and cells and
CC producing engineered breast tissue for treatment and research. The
CC transgenic animals are useful as animal model systems used in elaborating
CC the biological function of the polypeptides, studying conditions and/or
CC disorders associated with aberrant expression and in screening for
CC compounds effective in ameliorating the conditions. The polynucleotides
CC are useful for gene therapy and in vaccines
XX
SQ Sequence 1677 BP; 407 A; 456 C; 473 G; 341 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.24e-106 Length: 1677
Score: 1326.00 Matches: 260
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-078-090-151 (1-260) x ABV83605 (1-1677)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
DB 4 TTGGCGGTGCCAAGGCCACAGCGCTTGGGCTTCAGCATCCGTGGGGGCTCGGAGCAC 63

QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGlyLeu 40
DB 64 GCGGTGGGCATCTACGTGTCTCTGGTGGGAACCAAGGCTCTCTAGCTGAGAGGAGACTG 123

QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
DB 124 CGGGTGGGGACCAAGATTCTCGCGCTCAACGACAAATCCCTGGCCCGGTGACCCACGCG 183

QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
DB 184 GAGGCCGTCAAGGCTCTGAAGGGCTCCAAAGACTGGTGCTGTCTGTACTACAGCAGG 243

QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
DB 244 CGCATCCCTGGGGCTACGTACCAACCAACACATCTACACCTGGTGGACCCCGCAGGCGCG 303

QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
DB 304 AGCATCTCCCAACCTCGGGCTGCCCCAGCCCCACGGTGGTCCCTGAGGCAGCAGGAG 363

QY 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
DB 364 GGTGACCGGAGGAGCACCTGTCACCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTG 423

QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
DB 424 GTGCTGGGGGACGGCCGGTCCCTGGGCTCACGATCCGTGGGGGAGCTGAGTACGGCCTT 483

QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
DB 484 GGCATTTTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGCGGCGGCTCAAGGTT 543

QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
DB 544 GGGGACCAAGATCTAGAAAGTGAATGGGCGGAGCTTCTCAACATCTACACGAGGCT 603

QY 201 ValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220

DB 604 GTCAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTG 663

QY 221 ProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGlu 240
DB 664 CCCCATGCCCCGACCACTGTGGACGAGACCAAGTGGATCGCCAGTCCCGGATCAGGGAG 723

QY 241 ThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
DB 724 ACCATGGCGAACTCGCAGGGTCTGGCCACTCTGCTCGCTCCATCTCCAGACCCCGGG 783

RESULT 2
AAC75441
ID AAC75441 standard; cDNA; 3102 BP.
XX AAC75441;
AC AAC75441;
DT 08-FEB-2001 (first entry)
XX
DE Human ORF996 polynucleotide sequence SEQ ID NO:1991.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; SS.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41232.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 5; Page 1517-1519; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX

proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

Sequence 3102 BP; 699 A; 996 C; 860 G; 545 T; 0 U; 2 Other;

Alignment Scores:
 Pred. No.: 3,14e-73 Length: 3102
 Score: 954.00 Matches: 201
 Percent Similarity: 80.92% Conservative: 11
 Best Local Similarity: 76.72% Mismatches: 41
 Query Match: 71.95% Indels: 9
 DB: 3 Gaps: 3

US-10-078-090-151 (1-260) x AAC75441 (1-3102)

QY	1	LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis	20
Db	96	CTTCTGCAGAGAAATGCTTCTCAGGGGTTGGACTGTTCTGTAACCCCCCACTCCCGCC	155
QY	21	GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGlu-GlyLe	40
Db	156	AGCGCAGGTGTTTGAACCTCCAGCTGAGGCCTGCTGGCTGCTGGGAAACTCCTAGGCAG	215
QY	40	uArgValGlyAspGlnIleLeuArgValAlaAsnAspLysSerLeuAlaArgValThrHisAl	60
Db	216	CAGA--GGCCACGACTACTTCTCCTCAGTGCCGTTTCAGTGCC-----	258
QY	60	aGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGl	80
Db	259	----TGTGTCCAGGCTCTGAAGGGCTCCAAGAAGCTGGTGTCTGTGTACTCAGCAGG	314
QY	80	YArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyAr	100
Db	315	CGCATCCCTGGGGCTACGTCAACAACCATCTACACCTGGGTGGACCCGCGGCGG	374
QY	100	gSerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGl	120
Db	375	CAGCATCTCCCAACCTCGGGCCTGCCCAAGAGGGGATGAGAAAAAGGTGAACCT	434
QY	120	uGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLe	140
Db	435	GGGTGACCGGAGGAGACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAAAGGTGAACCT	499
QY	140	uValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLe	160
Db	495	GGTGCTGGGGACGCGCGGTCCCTGGGCCTCAGATCCGTGGGGAGCTGAGTACGGCCT	555
QY	160	uGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVa	180
Db	555	TGGCATTTACATCATCGCGTGGACCCAGGCTCTGAAGCAGAGGAGGAGGAGGAGGAGG	611
QY	180	lGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAl	200
Db	615	TGGGACCCAGATTCTAGAGTGAAATGGCGGAGCTTCTCAACATCTCTACAGCAGCAGGC	677
QY	200	aValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLe	220
Db	675	TGTCAGGCTGCTTAAGTCATCTCGGCACCTCATCTCTGACAGTGAGGAGCTCGGAGGCT	731
QY	220	uProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGl	240
Db	735	GCCCCATGCCCGCACCACTGTGGACGAGACCAAGTGGAATGCCAGATTCCCGGATCAGGGA	799
QY	240	uThrMetAlaAsnSerAlaGly--SerGlyHisSerAlaArgSerAsnLeuGlnThrPr	25
Db	795	GACCATGGCGAACTCGGCAGGTTTCTTGGCGATCTCACACAGAGGAATAACAAGCC	85

QY	259	oGly 260	
Db	855	AGGA 858	
RESULT 3			
ADA53214			
ID	ADA53214	standard; cDNA; 1718 BP.	
XX	AC		
AC	ADA53214;		
XX	20-NOV-2003	(first entry)	
XX	DT		
DE	Human coding sequence, SEQ ID 782.		
XX			
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;		
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;		
KW	inflammatory disease; osteoporosis; neurological disease; gene; ss.		
XX			
OS	Homo sapiens.		
XX	EP1293569-A2.		
PN			
XX	19-MAR-2003.		
PD			
XX			
PF	21-MAR-2002; 2002EP-00006586.		
XX			
PR	14-SEP-2001; 2001JP-00328381.		
PR	24-JAN-2002; 2002US-0350435P.		
XX			
PA	(HELI-) HELIX RES INST.		
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.		
XX			
PI	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;		
PI	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;		
PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;		
XX			
DR	WPI; 2003-395539/38.		
DR	P-PSDB; ADA54853.		
XX			
PT	New polynucleotides encoding full-length polypeptides, e.g. secretory		
PT	and/or membrane proteins, useful for developing medicines for diseases in		
PT	which the gene is involved, or as target molecules for gene therapy.		
XX			
PS	Claim 1; SEQ ID NO 782; 205pp; English.		
XX			
CC	The present invention relates to novel human secretory or membrane		
CC	proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-		
CC	ADA54071). The coding sequences are useful in the gene therapy of		
CC	diseases caused by abnormalities of the proteins, e.g. cancer,		
CC	inflammatory diseases, osteoporosis or neurological disease.		
XX			
SQ	Sequence 1718 BP; 269 A; 611 C; 551 G; 287 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	5.85e-52	Length:	1718
Score:	706.00	Matches:	140
Percent Similarity:	95.95%	Conservative:	2
Best Local Similarity:	94.59%	Mismatches:	6
Query Match:	53.24%	Indels:	0
DB:	7	Gaps:	0
US-10-078-090-151 (1-260) x ADA53214 (1-1718)			
QY	1	LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis	20
Db	1076	TTGCGGCGTGCCCAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC	113
QY	21	GlyValGlyIleTyValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu	40
Db	1136	GCGGTGGGCATCTACGTGTCTCTGGTGGAAACCAGGCTCTCTAGCTGAGAGGAAGGACTG	119
QY	41	ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla	60

Db 1196 CCGGTGGGGACCAAGATTCTGGCGTCAACGACAAATCCCTGGCCCGGTGACCCACGGG 1255
Qy 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 1256 GAGGCCGTCAAGGCTCTGAAGGGCTCCAAAGAACCTGGTGTCTGTGTACTCAGCAGGG 1315
Qy 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 1316 CGCATCTCTGGGGCTACGTCAACCAACCATCTACACCTGGTGGTGGACCCGACGGCCGC 1375
Qy 101 SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 1376 ACCATCTCCCAACCTCGGGCTGCCCCAGCCCGTGGTGGTGCCTGAGGCAGCAGGAG 1435
Qy 121 GlyAspArgArgSerThrIleuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
Db 1436 GGTGACCGGAGGAGCACCTGCACCTCTGCAGGAGGGGATGAGAAAAGGTGAGTGGG 1495
Qy 141 ValLeuGlyAspGlyArgSerLeu 148
Db 1496 GTGGGGAAGAGGAGCCAGCTCTC 1519

RESULT 4

ABX711178
ID ABX711178 standard; cDNA; 2822 BP.

XX
AC ABX711178;

DT 05-MAR-2003 (first entry)

XX
DE Novel human cDNA sequence #403.

XX Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
KW Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
KW haemostatic; antiinflammatory; expressed sequence tag; EST.

XX Homo sapiens.

XX WO200281731-A2.

XX 17-OCT-2002.

XX 29-JAN-2002; 2002WO-US001222.

XX 30-JAN-2001; 2001US-00774528.

XX (HYSE-) HYSEQ INC.

PA (GOOD/) GOODRICH R W.

XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;

XX WPI; 2003-058563/05.

XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid
PT or lymphoid cell disorders, bone disorders, mechanical and traumatic
PT disorders, coagulation disorders, and inflammatory diseases.

XX Claim 1; Page; 612pp; English.

CC This invention relates to the cDNA sequences encoding an isolated novel
CC human polypeptide. The protein encoded by the nucleic acid of the
CC invention is useful for treating central and peripheral nervous system
CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic

CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)
CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
CC bacterial, viral or fungal infections; allergic conditions such as
CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
CC disease, anaphylaxis). The protein may be used to inhibit the growth,
CC infection or function of infectious agents such as bacteria, fungi,
CC viruses, or to effect bodily characteristics, biorhythms or circadian
CC cycles of rhythms. The protein may also have
CC proliferation/differentiation, stem cell growth factor, haematopoiesis
CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
CC activities. The cDNA sequences of the invention are useful for expressing
CC recombinant protein for analysis. The present sequence represents a novel
CC human cDNA sequence of the invention, this sequence is an expressed
CC sequence tag (EST) and was identified using subtractive hybridisation
XX
SQ Sequence 2822 BP; 557 A; 893 C; 922 G; 450 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.26e-28 Length: 2822
Score: 438.00 Matches: 107
Percent Similarity: 55.13% Conservative: 38
Best Local Similarity: 40.68% Mismatches: 90
Query Match: 33.03% Indels: 28
DB: 7 Gaps: 6

US-10-078-090-151 (1-260) x ABX711178 (1-2822)

Qy 1 LeuArgArgAlaLysAlaHisGluGly--LeuGlyPheSerIleArgGlyGlySerGlu 19
Db 324 GTCCGGGTGGAGAGAGTCCAGCAGGGAGGCTGGGCTTCAGCGTGC CGGGGGCTCAGAG 383
Qy 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
Db 384 CATGGCTGGGCATCTTCGTGAGCAAGTGGAGGAGGAGCAGTGCGAGCGGSGTGGC 443
Qy 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
Db 444 CTGTGGTGGGGACAAGATCACGGAGGTGAATGGGTGAGCTGGAGAGCACCCACCATG 503
Qy 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
Db 504 GGTAGCGCCGTAAAGGTGCTGACCAGCAGCAGCGCGCTGCACATGATGTTCCGGCGCATG 563
Qy 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGly 99
Db 564 GGCCGTGTGCCGGGCATCAAGTTCTCCAAGGAGAGACACCGTGGTGGATGTGTGAAT 623
Qy 100 ArgSerIle-----SerProSerGlyLeuProGlnProHisGly 113
Db 624 CGCGGCTGGTAGTGAGAGAGTGGGTTCAACACCCCTCCGAC----- 665
Qy 114 GlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGly 133
Db 666 -----ACCAGCTCAGAAAGATGGTGTCCGGCGC---ATCGTCCACCTATACACAACCTCC 716
Qy 134 AspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArg 153
Db 717 GACGAC-----TTCTGCTGGGCTTCAACATCCGT 746
Qy 154 GlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAla 173
Db 747 GGGGGCAAGAGATTGGCCCTGGGCATCTATGTGTCCAAAGTGGACCATGTTGGGCTGGCC 806
Qy 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
Db 807 GAGGAGAAATGGCATCAAGGTGGGGACCAAGGTCTTGGCAGCCAAACGGTGTCAAGTTTGAC 866

QY 194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr 213
Db 867 GACATCAGCCACAGCCAGCGCGTGGAGGTGCTGAAGGCCAAACGACATCATGTGACC 926
QY 214 VallysAspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIle 233
Db 927 ATCAAGGAGACCGCGCGGTATCCTGCCTACAAGGAGATGGTTTCTGAGTACTGCTGGCTG 986
QY 234 -----AlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSer 251
Db 987 GACCGACTGAGCAACGGGTGCTGCAGCAGCTGTCTCCCGCCTCTGAGAGCAGTCCAGC 1046
QY 252 AlaArgSer 254
Db 1047 GTCTCTTCG 1055

RESULT 5
ABL29731
ID ABL29731 standard; DNA; 1338 BP.
XX
AC ABL29731;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40666.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.

XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

XX
PS Claim 1; SEQ ID NO 40666; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (AB857737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 1338 BP; 303 A; 400 C; 384 G; 251 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7.58e-19 Length: 1338
Score: 324.00 Matches: 78
Percent Similarity: 50.67% Conservative: 35
Best Local Similarity: 34.98% Mismatches: 54
Query Match: 24.43% Indels: 56
DB: 4 Gaps: 7

US-10-078-090-151 (1-260) x ABL29731 (1-1338)
QY 20 HisGlyValGly-----IleTyrValSerLeuValGluProGlySer 33
Db 46 CACGGCTTCGGTCCGCCCTCCCGGACCCCATGTACGCATCGATGGTCCCCCCC----- 99
QY 34 LeuAlaGluLysGluGlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLysSer 53
Db 100 -----CTGCACCCACAGAACCCAGCG 120
QY 54 LeuAlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuVal 73
Db 121 GCAGCG-----GCAGCAGCGGCGCGCGCTCCGGGGCG----- 156
QY 74 LeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThr 93
Db 157 -----GGTCTGCCCTTCCGCCACGACCTGCTCC 183
QY 94 TrpValAspProGlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGly 113
Db 184 TGGATGGACCGCCACCGACGACCCGCCCTCCCCG-----CCCATGGAGTACGGG 231
QY 114 GlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGly 133
Db 232 GGC-----AGGCGCAGCGAGCGAAGGATCGCATA----- 261
QY 134 AspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArg 153
Db 262 -----CGTCGCGTGGAGCTGCTGATCGAACCTGGTCACTCTCTGGCCCTGATGATCCGT 315
QY 154 GlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAla 173
Db 316 GCGCGGTGGAGTACGCGCTGGGATCTTTGTACCGCGGTAGATAAGGACAGCGTGGCA 375
QY 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
Db 376 GATCGATCCGGACTGATGATTGGCGACGAGATCCTCGAGGTCAATGGGCAATCCTTTCTC 435
QY 194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr 213
Db 436 GATGTGACGCACGACGAGGCGGTGGTTCAGTTGAAGTACCACAAACGCATGTGCTGGTG 495
QY 214 VallysAspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIle 233
Db 496 ATACGTGACGTGGGCAAGGTGCCCCACTCCTGCACCTCCATCGAGATGGAGCCCTGGGAC 555
QY 234 AlaSerSer 236
Db 556 GCCTACAGT 564

RESULT 6
ABX34636
ID ABX34636 standard; cDNA; 2259 BP.
XX
AC ABX34636;
XX
DT 13-FEB-2003 (first entry)
XX
DE Human mddt cDNA SEQ ID 197.
XX
KW MDDT; human; disease detection and treatment molecule polypeptide;
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
KW gene therapy; protein replacement therapy; cell proliferative disorder;
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW psoriasis; hepatitis; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200279449-A2.

XX PD 10-OCT-2002.

XX PF 27-MAR-2002; 2002WO-US009944.

XX PR 28-MAR-2001; 2001US-0279619P.

XX PR 29-MAR-2001; 2001US-0280067P.

XX PR 29-MAR-2001; 2001US-0280068P.

XX PR 16-MAY-2001; 2001US-0291280P.

XX PR 17-MAY-2001; 2001US-0291829P.

XX PR 17-MAY-2001; 2001US-0291849P.

XX PR 19-JUN-2001; 2001US-0299428P.

XX PR 20-JUN-2001; 2001US-0299776P.

XX PR 20-JUN-2001; 2001US-0300001P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;

PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX WPI; 2003-058431/05.

DR P-PSDB; ABU11646.

XX

PT New purified disease detection and treatment molecule proteins and

PT polynucleotides, useful for diagnosing; treating or preventing cancers

PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis

PT or hepatitis.

XX Claim 1; SEQ ID NO 197; 339pp + Sequence Listing; English.

PS

XX This invention describes a novel disease detection and treatment molecule

CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,

CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,

CC antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides

CC and the polypeptides of the invention can be used for gene therapy,

CC protein replacement therapy and are useful for treating a variety of

CC diseases or conditions. These polypeptides or polynucleotides are

CC particularly useful for diagnosing, treating or preventing cell

CC proliferative disorders (e.g. cancers including adenocarcinoma,

CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's

CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's

CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or

CC hepatitis. ABX34440-ABX34835 encode the MDDT polypeptides represented in

CC ABU11450-ABU11845, described in the disclosure of the invention. NOTE:

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 2259 BP; 555 A; 598 C; 661 G; 445 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.37e-18	Length:	2259
Score:	321.50	Matches:	94
Percent Similarity:	51.88%	Conservative:	44
Best Local Similarity:	35.34%	Mismatches:	102
Query Match:	24.25%	Indels:	27
DB:	7	Gaps:	8

US-10-078-090-151 (1-260) x ABX34636 (1-2259)

QY 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlyGlySerGlu 19

DB 355 GTGCGTCTGACCGTCTGCACCCCGAAGCCTCGGCCTGAGTGTGCTGGCTGGAG 414

QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39

DB 415 TTTGGCTGTGGCTCTTTCATCTCCACCTCATCAAGCGGTGTCAGCAGACAGCGTCGGG 474

QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59

Db 475 CTCCAGGTAGGGACGAGATCGTCGGATCAATGGATATTCCATCTCTCTCTACCCAT 534

QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79

Db 535 GAGGAGTCACTCAACCTCATTCGAACCAAGAAACTGTGTCCATCAAAAGTGAGACATC 594

QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro 97

Db 595 GGCCTGATCCCCGTGAAAAGCTCTCCTGATGAGCCCCCTCACTTGGCAGTATGTGGAT--- 651

QY 98 GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg 117

Db 652 -----CAGTTTGTGTGGAATCTGGGGCGTGGAGGCGAGCCTGGGC----- 693

QY 118 GlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLys 137

Db 694 ---TCCCCTGGAATCGGAAAC-----AAGGAGAAGAAG 726

QY 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157

Db 727 GTCTTCATCAGCCTGGTAGGCTCCCGAGGCCCTTGGCTGCAGCATTTCCAGCGGCCCATC 786

QY 158 TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly 177

Db 787 CAGAAGCCTGGCATCTTTATCAGCCATGTGAAACCTGGCTCCCTGTCTGTGCTGAGTGGGA 846

QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197

Db 847 TTGGAGATAGGGGACCAAGATTGTGGAAGTCAATGGCGTCGACTTCTTAACCTGGATCAC 906

QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr---ValLysAsp 216

Db 907 AAGGAGGCTGTAATGTGTGTAAGAGTAGCCGCGAGCCTGACCATCTCCATTGTAGTGCA 966

QY 217 ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer 236

Db 967 GCTGGCCGG-----GAGCTGTTTCATGACAGACCGGAGCGGCTGGCAGAGCGG 1014

QY 237 ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu 256

Db 1015 CGGCAGCGTGAGCT-GCAGCGCGCAGGAGCTTCTCATGCAAGAGCGGCTGGCGATGGAGTC 1073

QY 257 Gln-----ThrProGly 260

Db 1074 CAACAAGATCCTCCAGGA 1091

RESULT 7

AAAX40058

ID AAAX40058 standard; DNA; 2162 BP.

XX

AC AAAX40058;

XX

DT 02-JUL-1999 (first entry)

XX

DE Colon cancer associated gene.

XX

KW Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

KW prostate cancer; ss.

XX

OS Homo sapiens.

XX

PN WO9904265-A2.

XX

PD 28-JAN-1999.

XX

PF 15-JUL-1998; 98WO-US014679.

XX

PR 17-JUL-1997; 97US-00896164.

PR 10-OCT-1997; 97US-0061599P.

PR 10-OCT-1997; 97US-0061765P.

PR 10-OCT-1997; 97US-00948705.

PR 11-OCT-1997; 97GB-00021697.

CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer
XX
SO Sequence 2236 BP; 548 A; 588 C; 653 G; 447 T; 0 U; 0 Other;

Alignment Scores:		
Pred. No.:	4.27e-18	2236
Score:	318.50	93
Percent Similarity:	51.88%	45
Best Local Similarity:	34.96%	102
Query Match:	24.02%	27
DB:	2	8
	Length:	
	Matches:	
	Conservative:	
	Mismatches:	
	Indels:	
	Gaps:	

US-10-078-090-151 (1-260) x AAX40057 (1-2236)

QY	1	LeuArgArgAlaLysAlaHis---	GluGlyLeuGlyPheSerIleArgGlyGlySerGlu	19
DB	358	GTGCGTCTGGACCGTCTGCACCCCGAAGGCCTCGGCC	TGAGTGCCTGGTGGCCTGGAG	417
QY	20	HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly	39	
DB	418	TTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAAGGCGGTACGACGACAGCGTCCGG	477	
QY	40	LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis	59	
DB	478	CTCCAGGTAGGGACGAGATCGTCCGGATCAATGGATATTCCATCTCCTCTGTACCCAT	537	
QY	60	AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla	79	
DB	538	GAGGAGGTCATCAACCTCATTCGAACCAAGAAACTGTGTCCATCAAAAGTGAGACATC	597	
QY	80	GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro	97	
DB	598	GGCCTGATCCCGTGAAAAGCTCTCCTGATGAGCCCTCACTTGGCAGTATGTGGAT---	654	
QY	98	GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg	117	
DB	655	-----CAGTTTGTGCGAATCTGGGGCGGTGGCAGGCAGCCTGGGC-----	696	
QY	118	GlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLys	137	
DB	697	---TCCCCTGGAAATCGGGAAC-----AAGGAGAAGAAG	729	
QY	138	ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu	157	
DB	730	GTCTTCATCAGCCTGGTAGGCTCCCGAGGCCCTTGGTGCAGCATTTCCAGCGGCCCATC	789	
QY	158	TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly	177	
DB	790	CAGAAAGCCTGGCATCTTTATCAGCCATGTGAAACCTGGTCCCTGCTGCTGAGGTGGGA	849	
QY	178	LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis	197	
DB	850	TTGGAGATAGGGACCAAGATTGTGGAAGTCAATGGCGTCGACTTCTCTAACCTGGATCAC	909	
QY	198	AspGluAlaValArgLeuLeuLysSerSerArgHisIleIleLeuThr---ValLysAsp	216	
DB	910	AAGGAGGCTGTAAATGTGTCTGAAAAATAGCCGACGCTGACCATTCTCCATTGTAGCTGA	969	
QY	217	ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer	236	
DB	970	GCTGGCCGG-----GACCTGTTTCATGACAGACCCGGGAGCGGCTGGCAGAGCGG	1017	
QY	237	ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu	256	
DB	1018	CGGCAGCGTCAAGCT-GCAGCGGGCAGGAGCTTCTCATGCGAAGCGGGCTGGCGATGAGTC	1076	

QY	257	Gln-----ThrProGly 260
Db	1077	CAACAAGATCCTCCAGGA 1094
RESULT 9		
AA	AX	39644
ID	AA	AX39644 standard; DNA; 2236 BP.
XX	AC	AA
XX	AX	39644;
XX	DT	02-JUL-1999 (first entry)
XX	DE	Renal cancer associated gene.
XX	KW	Cancer associated antigen; diagnosis; research; treatment; human;
XX	KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX	KW	prostate cancer; ss.
XX	OS	Homo sapiens.
XX	PN	WO9904265-A2.
XX	PD	28-JAN-1999.
XX	PF	15-JUL-1998; 98WO-US014679.
XX	PR	17-JUL-1997; 97US-00896164.
XX	PR	10-OCT-1997; 97US-0061599P.
XX	PR	10-OCT-1997; 97US-0061765P.
XX	PR	10-OCT-1997; 97US-00948705.
XX	PR	11-OCT-1997; 97GB-00021697.
XX	PR	22-JUN-1998; 98US-00102322.
XX	PA	(LUDW-) LUDWIG INST CANCER RES.
XX	PI	Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
XX	PI	O'hare M, Obata Y, Pfreundschuh M, Tureci O, Sahin U;
XX	DR	WPI; 1999-132448/11.
XX	PT	New isolated cancer associated nucleic acids and polypeptides - isolated
XX	PT	using sera from cancer patients, used to develop products for the
XX	PT	diagnosis, monitoring or treatment of cancers.
XX	PS	Claim 67; Page 435-436; 787pp; English.

The invention relates to a method for diagnosing a disorder characterised by expression of a human cancer associated antigen precursor coded for by a nucleic acid molecule (NAM). The method comprises: (a) contacting a biological sample isolated from a subject with an agent that specifically binds to the NAM, an expression product or a fragment of an expression product complexed with an HLA molecule; and (b) determining the interaction between the agent and the NAM or the expression product as a determination of the disorder. The products and methods can be used in the diagnosis, monitoring, research, or treatment of conditions characterised by the expression of various cancer associated antigens. The invention provides nucleic acid sequences and encoded polypeptides which are cancer associated antigen precursors expressed in human breast cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and lung cancer

Sequence 2236 BP; 548 A; 588 C; 653 G; 447 T; 0 U; 0 Other

Alignment Scores:		
Pred. No.:	4.27e-18	2236
Score:	318.50	93
Percent Similarity:	51.88%	45
Best Local Similarity:	34.96%	102
Query Match:	24.02%	27
DB:	2	8
		Gaps:
		Indels:
		Mismatches:
		Conservative:
		Matches:
		Length:

US-10-078-090-151 (1-260) x AAX39644 (1-2236)

QY 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlyGlySerGlu 19
Db 358 GTGCGTCTGGACCGTCTGCACCCCGAAGCGCTCGGCCTGAGTGTGGCTGGAG 417
QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
Db 418 TTTGGCTGTGGGTCTTTCATCTCCACCTCATCAAAAGCGGTGAGGACAGACAGCGTGGG 477
QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
Db 478 CTCAGGTAGGGACGAGATCGTCCGGATCAATGGATATTCATCTCTCTCTCTCTCTCT 537
QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
Db 538 GAGGAGGTTCATCACTCATTCGAACCAAGAAACTGTGTCTCATCAAAAGTGAGACATC 597
QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro 97
Db 598 GGCCTGATCCCGCGTAAAGCTCTCTCTGATGAGCCCTCACTTGGCAGTATGTGGAT--- 654
QY 98 GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg 117
Db 655 -----CAGTTTGTGCGAATCTGGGGCGTGGAGGCGCTGGGC----- 696
QY 118 GlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLys 137
Db 697 ---TCCCTCGAATCGGAAAC-----AAGGAGAGAG 729
QY 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
Db 730 GTCTTCATCAGCCTGGTAGGCTCCCGAGGCTTGGCTGCAGATTTCCAGCGGCCCATC 789
QY 158 TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly 177
Db 790 CAGAGCCTGGCATCTTATCAGCCATGTGAACCTGGCTCCCTCTCTCTCTCTCTCTCTCT 849
QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
Db 850 TTGGAGATAGGGACAGATTTGCGAAGTCAATGGCGTGGCTGCTCTTAACCTGGATCAC 909
QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr---ValLysAsp 216
Db 910 AAGGAGCGTGTAAATGTCTGAAATAATAGCCGAGCCTGACCATCTCCATTTGTAGTGCA 969
QY 217 ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer 236
Db 970 GCTGGCGG-----GAGCTGTTCATGACAGACCGGAGCGGCTGGCAGAGCG 1017
QY 237 ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu 256
Db 1018 CGGCAGCGTGAGCT-GCAGCGGCGAGGCTTCTCATGCGAGAGCGGCTGGCGATGGAGTC 1076
QY 257 Gln-----ThrProGly 260
Db 1077 CAACAAGATCCTCCAGGA 1094
RESULT 10
ID AAX40061
XX AAX40061 standard; DNA; 2289 BP.
XX AAX40061;
XX 02-JUL-1999 (first entry)
XX Colon cancer associated gene.
DE Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX Homo sapiens.
OS WO9904265-A2.
PN

XX 28-JAN-1999.
XX 15-JUL-1998; 98WO-US014679.
XX 17-JUL-1997; 97US-00896164.
PR 10-OCT-1997; 97US-0061599P.
PR 10-OCT-1997; 97US-0061765P.
PR 10-OCT-1997; 97US-00948705.
PR 11-OCT-1997; 97GB-00021697.
PR 22-JUN-1998; 98US-00102322.
XX (LUDW-) LUDWIG INST CANCER RES.
XX Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
PI O'hare M, Obata Y, Pfreundschuh M, Tureci O, Sahin U;
XX WPI; 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides - isolated
PT using sera from cancer patients, used to develop products for the
PT diagnosis, monitoring or treatment of cancers.
XX Claim 67; Page 663-664; 787pp; English.
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer
XX
SQ Sequence 2289 BP; 567 A; 596 C; 668 G; 458 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 4.4e-18 Length: 2289
Score: 318.50 Matches: 93
Percent Similarity: 51.88% Conservative: 45
Best Local Similarity: 34.96% Mismatches: 102
Query Match: 24.02% Indels: 27
DB: 2 Gaps: 8
US-10-078-090-151 (1-260) x AAX40061 (1-2289)
QY 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlyGlySerGlu 19
Db 358 GTGCGTCTGGACCGTCTGCACCCCGAAGCGCTCGGCCTGAGTGTGGCTGGAG 417
QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
Db 418 TTTGGCTGTGGGTCTTTCATCTCCACCTCATCAAAAGCGGTGAGGACAGACAGCGTGGG 477
QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
Db 478 CTCAGGTAGGGACGAGATCGTCCGGATCAATGGATATTCATCTCTCTCTCTCTCTCT 537
QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
Db 538 GAGGAGGTTCATCACTCATTCGAACCAAGAAACTGTGTCTCATCAAAAGTGAGACATC 597
QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro 97
Db 598 GGCCTGATCCCGCGTAAAGCTCTCTCTGATGAGCCCTCACTTGGCAGTATGTGGAT--- 654
QY 98 GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg 117

D	b	655	----	---CAGTTTGTGTCGGAATCTGGGGCGTGCCGAGGCAGCCTGGGC-----			696
Q	y	118	GlnGlnGluGlyAspArgSerThrLeuHisLeuGlnGlyGlyAspGluLysLys				137
D	b	697	---	TCCCCTGGAAATCGGAAAAC-----		AAGGAGAAGAAG	729
Q	y	138	ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu				157
D	b	730	GTCTTCAATCAGCTGGTAGGCTCCCGAGGCCCTGGCTGCAGCATTTCCAGCGCCCCATC				789
Q	y	158	TyrGlyLeuGlyIleTyriLeThrGlyValAspProGlySerGluAlaGluGlySerGly				177
D	b	790	CAGAAAGCCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGTCTGTGAGGTGGGA				849
Q	y	178	LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis				197
D	b	850	TTGGAGATAGGGACCAGATTGTCGAAGTCAATGGCGTCGACTTCTCTAACCTGGATCAC				909
Q	y	198	AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr--ValllysAsp				216
D	b	910	AAGGAGGCTGTAAATGTGCTGAAAAAATAGCCGACGCTGACCATCTCCATTGTAGCTGCA				969
Q	y	217	ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer				236
D	b	970	GCTGGCCGG-----GAGCTGTTTCATGACAGACCGGAGCGGCTGGCAGAGGCG				1017
Q	y	237	ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu				256
D	b	1018	CGGCAGCGTGAGCT-GCAGCGGCAGGAGCTTCTCATGCAGAAGCGGCTGGCGATGGAGTC				1076
Q	y	257	Gln-----ThrProGly	260			
D	b	1077	CAACAAGATCCTCCAGGA	1094			
 RESULT 11 AAX40062							
ID		AAX40062 standard; DNA; 2409 BP.					
XX							
AC		AAX40062;					
XX							
DT		02-JUL-1999 (first entry)					
XX							
DE		Colon cancer associated gene.					
XX							
KW		Cancer associated antigen; diagnosis; research; treatment; human;					
KW		breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;					
KW		prostate cancer; ss.					
XX							
OS		Homo sapiens.					
XX							
PN		WO9904265-A2.					
XX							
XX		28-JAN-1999.					
XX							
Pf		15-JUL-1998; 98WO-US014679.					
XX							
PR		17-JUL-1997; 97US-00896164.					
PR		10-OCT-1997; 97US-0061599P.					
PR		10-OCT-1997; 97US-0061765P.					
PR		10-OCT-1997; 97US-0094870S.					
PR		11-OCT-1997; 97GB-00021697.					
PR		22-JUN-1998; 98US-00102322.					
XX							
PA		(LUDW-) LUDWIG INST CANCER RES.					
XX							
PI		Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;					
PI		O'hare M, Obata Y, Pfreundschuh M, Tureci O, Sahin U;					
XX							
DR		WPI; 1999-132448/11.					
XX							
PT		New isolated cancer associated nucleic acids and polypeptides - isolated					
PT		using sera from cancer patients, used to develop products for the					

diagnosis, monitoring or treatment of cancers.

Claim 67; Page 665-666; 787pp; English.

The invention relates to a method for diagnosing a disorder characterised by expression of a human cancer associated antigen precursor coded for by a nucleic acid molecule (NAM). The method comprises: (a) contacting a biological sample isolated from a subject with an agent that specifically binds to the NAM, an expression product or a fragment of an expression product complexed with an HLA molecule; and (b) determining the interaction between the agent and the NAM or the expression product as a determination of the disorder. The products and methods can be used in the diagnosis, monitoring, research, or treatment of conditions characterised by the expression of various cancer associated antigens. The invention provides nucleic acid sequences and encoded polypeptides which are cancer associated antigen precursors expressed in human breast cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and lung cancer

Sequence 2409 BP; 595 A; 638 C; 692 G; 484 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.68e-18	Length:	2409
Score:	318.50	Matches:	93
Percent Similarity:	51.88%	Conservative:	45
Best Local Similarity:	34.96%	Mismatches:	102
Query Match:	24.02%	Indels:	27
DB:	2	Gaps:	8

US-10-078-090-151 (1-260) x AAX40062 (1-2409)

QY	1	LeuArgArgAlaLysAlaHis--GluGlyLeuGlyPheSerIleArgGlyGlySerGlu	19
DB	358	GTGCGTCTGGACCGTCTGCACCCCGAAGGCCTCGGCCTGAGTGTGCGTGGTGGCTGGAG	417
QY	20	HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly	39
DB	418	TTTGGCTGTGGGTCTTTCATCTCCACCTCATCAAAAGCGGTGAGGCAGACGCGTGGG	477
QY	40	LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis	59
DB	478	CTCCAGTAGGGACGAGATCGTCCGGATCAATGGATATTCCATCTCCTCTGTACCCAT	537
QY	60	AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla	79
DB	538	GAGGAGTCAACAACCTCATTCGAACCAAGAAACTGTGTCCATCAAAGTGAGACATC	597
QY	80	GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro	97
DB	598	GGCCTGATCCCGTGAAAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTGGAT---	654
QY	98	GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg	117
DB	655	-----CAGTTTGTGTCGGAATCTGGGGCGGTGCGAGCAGCCTGGGC-----	696
QY	118	GlnGlnGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLys	137
DB	697	---TCCCTGGAATCGGAAAC-----AAGGAGAAGAAG	729
QY	138	ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu	157
DB	730	GTCTTCATCAGCCTGGTAGGCTCCCGAGGCCTTGGCTGCAGCATTTCCAGCGGCCCATC	789
QY	158	TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly	177
DB	790	CAGAAGCCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGTCTGTGAGGTGGGA	849
QY	178	LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis	197
DB	850	TTGGAGATAGGGACCAGATTGTCGAAGTCAATGGCGTCGACTCTCTAACTGGATCAC	909
QY	198	AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr---ValLysAsp	216

Db 910 AAGGAGGCTGTAATGTCTGAAATAAGCCGAGCCTGACCATCTCCATTGTAGCTGCA 969
Qy 217 ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer 236
Db 970 GCTGGCCGG-----GAGCTGTTTCATGACAGACCGGAGCGGCTGGCAGAGCG 1017
Qy 237 ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu 256
Db 1018 CGCAGCGTGAGCT-GCAGCGCAGGAGCTTCTCATGTCAGAAAGCGGCTGGCGATGGAGTC 1076
Qy 257 Gln-----ThrProGly 260
Db 1077 CAACAAGATCCTCCAGGA 1094

RESULT 12
ABZ75586
ID ABZ75586 standard; DNA; 837 BP.
XX AC ABZ75586;
XX DT 28-APR-2003 (first entry)
XX DE SID polypeptide encoding DNA.
XX KW SID; deafness; hearing disorder; auditory; gene therapy; human; ds.
XX OS Unidentified.

Key Location/Qualifiers
FH 1..837
FT /*tag= a
FT /product= "SID"
XX WO2003000891-A2.
XX PN 03-JAN-2003.
XX PD 21-JUN-2002; 2002WO-EP008065.
XX PF 21-JUN-2001; 2001US-0299848P.
XX PR 05-FEB-2002; 2002EP-00290277.
XX (HYBR-) HYBRIGENICS.
PA (INSP) INST PASTEUR.

Daviet L, Legrain P, Petit C, Boeada B, El-Amraoui A;
WPI; 2003-201427/19.
DR P-PSDB; ABP71383.

New complexes of Selected Interacting Domains of polypeptides or
polynucleotides, useful for preventing and treating deafness and other
inner ear or hearing disorders in humans or animals.

Claim 13; Page 62-63; 92pp; English.
The invention relates to protein-protein interactions involved in
deafness or hearing disorders, specifically proteins that interact with
mammalian inner ear proteins. Complexes of selected interacting domains
(SID) of polypeptides are useful in preventing and treating deafness and
other inner ear or hearing disorders and/or diseases in humans or
animals. The methods are useful in screening drugs for agents which
modulate protein interaction. The present sequence represents a DNA
encoding the SID polypeptide that interacts with the human myosin VIIA

Sequence 837 BP; 201 A; 211 C; 270 G; 155 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.74e-18 Length: 837
Score: 312.00 Matches: 83
Percent Similarity: 51.53% Conservative: 35
Best Local Similarity: 36.24% Mismatches: 83
Query Match: 23.53% Indels: 28

DB: 7 Gaps: 6
US-10-078-090-151 (1-260) x ABZ75586 (1-837)
Qy 8 GluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyValGlyIleTyrValSer 27
Db 19 GAAGGTCTCGGCTCAGCGTCGCTGGAGGCTTGAATTTGGCTGTGGACTCTTTATCTCC 78
Qy 28 LeuValGluProGlySerLeuAlaGluLysGluGlyLeuArgValGlyAspGlnIleLeu 47
Db 79 CACCTCATCAAAGGTGGCCAGGCAGACAGCGTTGGCTTCAGGTAGGGGATGAATGTC 138
Qy 48 ArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAlaValLysAlaLeuLys 67
Db 139 CGGATCAACGGCTATTCCATCTCTTCTGTACCCATGAGGAAGTCATCAACCTGATCCGC 198
Qy 68 GlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrVal 87
Db 199 ACCAAGAAGACCGTGTCCATCAAAAGTGAGACACATCGGACTGATCCCTGTGAAGAGCTCT 258
Qy 88 ThrAsnHisIleTyrThrTrp-----ValAspProGlnGlyArgSerIleSerProPro 105
Db 259 CCTGAGGAGTCCCTCAAATGGCAGTATGTGGAT-----CAGTTCTGTCGGAATCT 309
Qy 106 SerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGluGlyAspArgArgSer 125
Db 310 GGGGTCTG-----CGAGGTGGCTTG-----GGCTCACCTGGCAATCGGACACACC 354
Qy 126 ThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsnLeuValLeuGlyAspGly 145
Db 355 -----AAGGAGAAGAAGGTGTTTATCAGTCTAGTGGGCTCT 390
Qy 146 ArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThr 165
Db 391 CGGGGCTGGGCTGCAGCATCTCCAGTGGCCCCCATCCAGAACCTGGCATCTTCGTCAAGC 450
Qy 166 GlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIleLeu 185
Db 451 CACGTGAAGCCTGGCTCCCTGTCTGCAGAGGTGGGGTTAGAGACAGGACCAGATTGTG 510
Qy 186 GluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLeuLys 205
Db 511 GAAGTCAATGGCATAGACTTCACCAACCTGGACCACAAAGGAGGCTGTGAATGTCTGAAG 570
Qy 206 SerSerArgHisLeuIleLeuThrVal-----Lys 215
Db 571 AGCAGCCGAGCCTGACCATCTCCATCGTTGTGGAGCCGCGGAGCTGTTTCATGACG 630
Qy 216 AspValGlyArgLeuProHisAlaArg 224
Db 631 GACCGGGAACGCGTGGAGGAGGCACGG 657

RESULT 13
ABL29730/c
ID ABL29730 standard; DNA; 6497 BP.
XX AC ABL29730;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40663.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.

QY 131 GlnGlyGlyAspGluLysLysValAsnLeuValLeu-----GlyAspGlyArgSerLeu 148
Db 574 -----GGTGGACCGCGGACATCAGTGTGGTCTCTCCATGTGGCGCCAGGACGAAGCTG 627
QY 149 GlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAsp 168
Db 628 GGACTGGGATTTCCAAGGACCCGAGTGGAAAGCCGGGCATCTTCGTGCAGTTCCACAAAG 687
QY 169 ProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsn 188
Db 688 GATCGGAGTGTGCGCGGAGCGGGATTACGGCCAGGTGACCAAGATCCTCAGCGTCAAC 747
QY 189 GlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArg 208
Db 748 AGCATCGACTTCTCGGACGTCTCTTTAGCGAGCGCGTGGCCGTGATGAAGAGCAGCAGC 807
QY 209 HisLeuIleLeuThrValLys 215
Db 808 AAGCTGGACATGGTGTGCGC 828
RESULT 15
ACC85071
ID ACC85071 standard; DNA; 3046 BP.
XX
AC ACC85071;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human MBCAT polypeptide encoding DNA.
KW MBCAT; beta-catenin; cytostatic; gene therapy; cancer; human; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 189..2969
FT /*tag= a
FT /product= "MBCAT"
XX
PN WO2003052068-A2.
XX
PD 26-JUN-2003.
XX
PF 12-DEC-2002; 2002WO-US039796.
XX
PR 13-DEC-2001; 2001US-0340213P.
PR 13-DEC-2001; 2001US-0340314P.
PR 13-DEC-2001; 2001US-0340322P.
PR 15-FEB-2002; 2002US-0357502P.
XX
PA (EXEL-) EXELIXIS INC.
XX
PI Costa MA, Gendreau SB, Dora EG, Nicoll M;
XX
DR WPI; 2003-533010/50.
DR P-PSDB; ABR82220.
XX
PT Identifying a candidate beta-catenin pathway modulating agent for
PT diagnosing or treating cancer by detecting a test agent-biased activity
PT of the assay system comprising a purified MBCAT polypeptide or nucleic
PT acid.
XX
PS Example; Page 49-51; 81pp; English.
XX
CC The invention relates to genes that modify beta-catenin pathway and to
CC the identification of human MBCAT (modifiers of beta-catenin) genes. The
CC MBCAT polypeptides are therapeutic targets for disorders associated with
CC beta-catenin function and are useful for manufacturing a medicament for
CC diagnosing or treating breast, colon, lung or ovary cancer. The present
CC sequence represents a human MBCAT polypeptide encoding DNA
XX
SQ Sequence 3046 BP; 975 A; 612 C; 681 G; 778 T; 0 U; 0 Other;

Alignment Scores: 2.57e-10 Length: 3046
Pred. No.: 231.00 Matches: 91
Score: 46.91% Conservative: 38
Percent Similarity: 33.09% Mismatches: 100
Best Local Similarity: 17.42% Indels: 46
Query Match: 8 Gaps: 15
DB:
US-10-078-090-151 (1-260) x ACC85071 (1-3046)
QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGly-----GlySer 18
Db 1146 ATAAAGCTCATTTAAAGGTCCTAAAGGTCITGGGTTTAGCATTTGCTGGAGGTGTTGAAAT 1205
QY 19 GluHisGlyValGly-----IleTyrValSerLeuValGluProGlySerLeuAla 35
Db 1206 CAGCATATTCTCTGGGATAATAGCATCTATGTAAACCAAAATAATTGAAGGAGGTGCAGCA 1265
QY 36 GluLysGluGly----LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeu 54
Db 1266 CATAAGGATGGCAAACTTCAGATTGGAGATAAACTTTTAGCAGTGAATAACGTATGTTTA 1325
QY 55 AlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuVal--- 73
Db 1326 GAAGAAGTTACTCATGAAGAAGCAGTAACTCCTTAAAGAACACATCTCATTTTGTATTAT 1385
QY 74 -----LeuSerValTyrSerAlaGlyArgIleProGlyGlyTyr--- 86
Db 1386 TTGAAAGTGGCAAAACCCACAAGTATGTAT-----ATGAATGATGGCTATGCA 1433
QY 87 -----ValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIleSer 103
Db 1434 CCACCTGATATCACCAAC-----TCCTCTCTCAGCCTGTTGTATAACCATGTTAGC 1484
QY 104 ProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAspArg 123
Db 1485 CCATCTTCTTCTTGGGCCAGACACACAGCATCTCCAGCCAGATACTCCCGAGTTTCTAAA 1544
QY 124 ArgSerThrLeuHisLeuLeuGlnGlyGlyAspGlu-----LysLysVal 138
Db 1545 -----GCAGTACTTGGAGATGATGAAATTTACAAGGGAAACCTAGAAAAAGTT 1589
QY 139 AsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyr 158
Db 1590 GTTCTTCATCGTGGC--TCAACGGGCGCTTGGTTTCAACATTGTAGGAGGA---GAAGAT 1643
QY 159 GlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly--- 177
Db 1644 GGAGAAGGAATATTATTTCCTTTATCTTAGCCGGAGGACCTGCTGATCTAAGTGGAGAG 1703
QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
Db 1704 CTCAGAAAAGGAGATCGTATTATATCGGTAACAGTGTGTGACCTCAGAGCTGTAGTCAT 1763
QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspVal 217
Db 1764 GAGCAGGCAGCAGCTGCATTGAAAATGCTGGCCAGGCTGTCACAATTGTGCACAATAT 1823
QY 218 GlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArg 237
Db 1824 CGACCTGAAGAAATACAGTCGTTTT-----GAAGCTAAA-----ATACATGAT 1865
QY 238 IleArgGluThrMetAlaAsn-----SerAlaGlySerGly 249
Db 1866 TTACGGGAGCAGATGATGAATAGTAGTATTAGTTCAGGGTTCAGGT 1910

Search completed: April 22, 2004, 02:06:17
Job time : 381 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 22, 2004, 01:43:37 ; Search time 388 Seconds
(without alignments)
3021.245 Million cell updates/sec

Title: US-10-078-090-151
Perfect score: 1326
Sequence: 1 LRRKAHEGLGFSIRGSEH.....TWANSAGSGHSARSNLQTPG 260

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2907579 seqs, 2254313464 residues
Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10078090/runat_20042004_132811_16613/app query.fasta_1.455
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10078090@cgn_1_1_333_@runat_20042004_132811_16613
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

1	1326	100.0	1677	15	US-10-078-090-48	Sequence 48, Appl
2	951.5	71.8	2964	13	US-10-112-944-536	Sequence 536, App
3	706	53.2	1718	16	US-10-094-749-782	Sequence 782, App
4	438	33.0	2822	16	US-10-120-988-406	Sequence 406, App
5	408.5	30.8	509	15	US-10-029-386-3728	Sequence 3728, Ap
6	346	26.1	195	15	US-10-029-386-17428	Sequence 17428, A
7	312	23.5	837	15	US-10-177-191A-5	Sequence 5, Appli
8	229	17.3	3071	16	US-10-305-720-1138	Sequence 1138, Ap
9	226.5	17.1	5139	13	US-10-451-207-14	Sequence 14, Appl
10	222	16.7	3995	9	US-09-919-497-9	Sequence 9, Appli
11	222	16.7	3995	13	US-10-170-385-300	Sequence 300, App
12	203	15.3	4235	13	US-10-276-774-724	Sequence 724, App
13	203	15.3	4235	13	US-10-296-115-496	Sequence 496, App
14	198	14.9	8640	13	US-09-964-956-24	Sequence 24, Appl
15	198	14.9	8640	13	US-09-964-956-26	Sequence 26, Appl
16	197.5	14.9	5836	10	US-09-998-425-2	Sequence 2, Appli
17	197.5	14.9	5836	10	US-09-997-977-2	Sequence 2, Appli
18	197	14.9	7516	9	US-09-917-800A-1664	Sequence 1664, Ap
19	196.5	14.8	2000	13	US-10-342-887-1009	Sequence 1009, Ap
20	196.5	14.8	2000	13	US-10-172-118-1009	Sequence 1009, Ap
21	196.5	14.8	2000	15	US-10-177-293-427	Sequence 427, App
22	196.5	14.8	2411	9	US-09-823-356-23	Sequence 23, Appl
23	196	14.8	1578	9	US-09-880-107-3440	Sequence 3440, Ap
24	195	14.7	1750	15	US-10-210-120-80	Sequence 80, Appl
25	195	14.7	3180	16	US-10-104-047-796	Sequence 796, App
26	195	14.7	6582	13	US-10-342-887-958	Sequence 958, App
27	195	14.7	6582	13	US-10-172-118-958	Sequence 958, App
28	195	14.7	7431	9	US-09-951-401-2	Sequence 2, Appli
29	195	14.7	7431	9	US-09-922-101-2	Sequence 2, Appli
30	195	14.7	7431	9	US-09-951-402-2	Sequence 71, Appl
31	194	14.6	3259	14	US-10-001-843-71	Sequence 68, Appl
32	191	14.4	6383	9	US-09-728-952-68	Sequence 1545, Ap
33	189.5	14.3	1152	15	US-10-106-698-1545	Sequence 75, Appl
34	189	14.3	3170	13	US-10-114-270-75	Sequence 79, Appl
35	186.5	14.1	1905	13	US-10-210-281-79	Sequence 27, Appl
36	186.5	14.1	3520	13	US-10-362-892-27	Sequence 27, Appl
37	186.5	14.1	3520	16	US-10-288-798-27	Sequence 9, Appli
38	186.5	14.1	5640	15	US-10-393-892-9	Sequence 9, Appli
39	186.5	14.1	5640	15	US-10-394-382-9	Sequence 99, Appl
40	186.5	14.1	6588	13	US-10-112-944-99	Sequence 1259, Ap
41	186.5	14.1	8062	13	US-10-342-887-1259	Sequence 1259, Ap
42	186.5	14.1	8062	13	US-10-172-118-1259	Sequence 45, Appl
43	186.5	14.1	8119	9	US-09-802-669-45	Sequence 45, Appl
44	186.5	14.1	8119	13	US-10-619-220-45	Sequence 45, Appl
45	183	13.8	3885	10	US-09-836-499-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-10-078-090-48
; Sequence 48, Application US/10078090
; Publication No. US20030044815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 1677

TYPE: DNA
ORGANISM: Homo sapien
US-10-078-090-48

Alignment Scores:
Pred. No.: 2,74e-145 Length: 1677
Score: 1326.00 Matches: 260
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-078-090-151 (1-260) x US-10-078-090-48 (1-1677)

Qy 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 4 TTGGCGGCTGCCAAGGCCCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 63
Qy 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGlyLeu 40
Db 64 GGCCTGGGCATCTACGTGTCTCTGGTGGAACCCAGGCTCTAGCTGAGAAGGAGGACTG 123
Qy 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 124 CGGCTCGGGGACCCAGATTCTTGGCGGTCAACGACAAATCCCTGGCCCGGTGACCCACGCG 183
Qy 61 GluAlaValLysAlaLeuLysGlySerLysLeuValLeuSerValTyrSerAlaGly 80
Db 184 GAGGCCGTCAAGGCTCTGAAGGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGG 243
Qy 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 244 CGCATCTCCCGCCACCTCGGGCTGACGTCAACCAACCATCTACACCTGGTGGACCCCGCAGGCCGC 303
Qy 101 SerIleSerProProSerGlyLeuProGlnProHisGlyAlaLeuArgGlnGlnGlu 120
Db 304 AGCATCTCCCGCCACCTCGGGCTGACGTCAACCAACCATCTACACCTGGTGGACCCCGCAGGAG 363
Qy 121 GlyAspArgSerThrLeuHisLeuLeuGlnGlyAspGluLysLysValAsnLeu 140
Db 364 GGTGACCGGAGAGCACCTTGACCTCTGCAAGGAGGGGATGAGAAAAGGTGACCTG 423
Qy 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyAlaGluTyrGlyLeu 160
Db 424 GTGCTGGGGGACGGCCGCTCCCTGGGCTCAGCATCCGTGGGGAGCTGAGTACGGCCTT 483
Qy 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
Db 484 GGCATTTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGCGCGGCTCAAGGTT 543
Qy 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
Db 544 GGGGACCGAGATTCTAGAAGTGAATGGCGGAGCTTTCTCAACATCTTACACGACGAGGCT 603
Qy 201 ValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
Db 604 GTCAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGACGTGGGAGGCTG 663
Qy 221 ProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGlu 240
Db 664 CCCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTCCCGGATCAGGGAG 723
Qy 241 ThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
Db 724 ACCATGGCGAACTCGGACGGGTCTGGCCACTCTGCTCGTCCAATCTCCAGACCCCGAGG 783

RESULT 2

US-10-112-944-536
Sequence 536, Application US/10112944
Publication No. US20040048249A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Yang, Yonghong

APPLICANT: Weng, Gezhi
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aidong J.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Dunrui
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Zhiwei
TITLE OF INVENTION: No. US20040048249A1el Nucleic Acids and
SECRETED POLYPEPTIDES
FILE REFERENCE: 805A
CURRENT APPLICATION NUMBER: US/10/112,944
CURRENT FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: US 09/519,705
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: US 09/552,929
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 09/577,408
PRIOR FILING DATE: 2000-05-18
NUMBER OF SEQ ID NOS: 924
SOFTWARE: pt_FL_genes Version 5.0
SEQ ID NO 536
LENGTH: 2964
TYPE: DNA
ORGANISM: Homo sapiens
US-10-112-944-536

Alignment Scores:

Pred. No.: 4,79e-101 Length: 2964
Score: 951.50 Matches: 191
Percent Similarity: 89.95% Conservative: 6
Best Local Similarity: 87.21% Mismatches: 15
Query Match: 71.76% Indels: 7
DB: 13 Gaps: 2

US-10-078-090-151 (1-260) x US-10-112-944-536 (1-2964)

Qy 43 GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
Db 20 GGGCCACGACTACTCTCTCCTGAGTGGCTTCACTGGCC-----TGT 61
Qy 63 ValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIle 82
Db 62 GTCCAGGCTCTGAAGGGCTCCAAGAGCTGGTGTGTGTGTACTCAGCAGGCGCATC 121
Qy 83 ProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIle 102
Db 122 CCTGGGGCTACGTCAACCAACCATCTACACCTGGTGGACCCCGCAGGCGCAGCATC 181
Qy 103 SerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAsp 122
Db 182 TCCCCACCCCTCGGGCTGCCCGCCAGCCCGGCTGGTGGTGGTGGTGGTGGTGGTGGT 241
Qy 123 ArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeuValLeu 142
Db 242 CGGAGGAGCACCTGTCACCTCTCTCAAGGAGGGGATGAGAAAAGGTGAACCTGGTGTG 301
Qy 143 GlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIle 162
Db 302 GGGGACGGCCGCTCCCTGGGCTCAGCATCCGTGGGGGAGCTGAGTACGGCCTTGGCATT 361


```
QY 163 TyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAsp 182
Db 362 TACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAAAGGACGGGCTCAAGGTTGGGGAC 421
QY 183 GlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArg 202
Db 422 CAGATTCTAGAAAGTGAATGGCGGAGCTTCTCAACATCCTACACGACGAGGCTGTCAAG 481
QY 203 LeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeuProHis 222
Db 482 CTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCCGGAGGCTGCCCAT 541
QY 223 AlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMet 242
Db 542 GCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCCCGGATCAGGGAGACCATG 601
QY 243 AlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
Db 602 GCGAACTCGGCAGGTTTCTTGGCGATCTCACAACAGAAGGAATAAACAAGCCAGGA 658
RESULT 3
US-10-094-749-782
; Sequence 782, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 782
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-782
Alignment Scores:
Pred. No.: 1.38e-72 Length: 1718
Score: 706.00 Matches: 140
Percent Similarity: 95.95% Conservative: 2
Best Local Similarity: 94.59% Mismatches: 6
Query Match: 53.24% Indels: 0
DB: 16 Gaps: 0
```

```
US-10-078-090-151 (1-260) x US-10-094-749-782 (1-1718)
QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 1076 TTGCGGCGTGCCCAAGGCCACGAGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 1135
```

```
QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 1136 GGCCTGGGCATCTACGTGTCTCTGGTGAACACGAGCTCTCTAGCTGAGAAGGAGGACTG 1195
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 1196 CGGGTCGGGGACCAAGATTCTCGCGCTCAACGACAAATCCCTGGCCCGGTGACCCACGCG 1255
QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 1256 GAGGCCGTCAAGGCTCTGAAGGCTCCAAGAAGCTGGTGTCTGTGTACTCAGCAGGG 1315
QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 1316 CGCATCTCTGGGGGTACGTCAACCAACCATCTACACCTGGGTGGACCCCGCAGGCCGC 1375
QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 1376 AGCATCTCCCCACCCCTCGGGCTGCCCCAGCCCGACGGTGGTGGCTGAGGCAGGAG 1435
QY 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
Db 1436 GGTGACCGGAGGAGCACCCCTGCACCTCTCTCAAGGAGGGATGAGAAAAAGGTGAGTGGG 1495
QY 141 ValLeuGlyAspGlyArgSerLeu 148
Db 1496 GTCGGGAAGGAGGCCAGCCTCTC 1519
RESULT 4
US-10-120-988-406
; Sequence 406, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyan
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 802CON
; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 406
; LENGTH: 2822
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (66)..(2822)
US-10-120-988-406
Alignment Scores:
Pred. No.: 7.15e-41 Length: 2822
Score: 438.00 Matches: 107
Percent Similarity: 55.13% Conservative: 38
Best Local Similarity: 40.68% Mismatches: 90
Query Match: 33.03% Indels: 28
DB: 16 Gaps: 6
```

```
US-10-078-090-151 (1-260) x US-10-120-988-406 (1-2822)
QY 1 LeuArgArgAlaLysAlaHisGluGly---LeuGlyPheSerIleArgGlyGlySerGlu 19
Db 324 GTCCGGGTGGAGAAGAGTCCAGCAGGAGGCTGGGCTTCAGCGTCCCGGGGCTCAGAG 383
QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
```

Db 384 CATGGCCTGGGCATCTTCGTCAGCAAAAGTGGAGGAAGGCAGCAGTGCGAGCGGGCTGGC 443
QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
Db 444 CTGTGCGTGGGGACAAGATCACGGAGGTGAATGGGCTGAGCCTGGAGAGCACCAACCATG 503
QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
Db 504 GGTAGCGCCGTAAGGTGCTGACCAAGCAGCAGCGCGCTGCACATGATGGTTCGGCGCATG 563
QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGly 99
Db 564 GGCCTGTGCCGGCATCAAGTCTCTCCAAGGAGAGACCAACCGTGGTGGATGTGTAAT 623
QY 100 ArgSerIle-----SerProProSerGlyLeuProGlnProHisGly 113
Db 624 CGGCCCTGGTAGTGGAAGAGTGGGTTCAACACCCCTCCGAC----- 665
QY 114 GlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGly 133
Db 666 -----ACCAGCTCAGAAGATGGTGTCCGGCGC---ATCGTCCACCTATACACAACCTCC 716
QY 134 AspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArg 153
Db 717 GACGAC-----TTCTGCTGGGCTTCAACATCCGT 746
QY 154 GlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAla 173
Db 747 GGGGGCAAGAGTTTGGCCTGGGCATCTATGTCTCCAAAGTGGACCATGGTGGGCTGGCC 806
QY 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
Db 807 GAGGAGATGGCATCAAGGTGGGGACCAAGGTCCTGGCAGCCACGGTGTCAAGTTTGAC 866
QY 194 AsnIleLeuHisAspGluAlaValArgLeuLysSerSerArgHisLeuIleLeuThr 213
Db 867 GACATCAGCCACAGCCAGGCCGTGGAGGTCTGAAGGGCCAAACGCACATCATGCTGACC 926
QY 214 ValLysAspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIle 233
Db 927 ATCAAGGAGACCGCCGGTATCCTGCTACAAAGAGATGGTTTCTGAGTACTGCTGGCTG 986
QY 234 -----AlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSer 251
Db 987 GACCGACTGAGCAACGGGGTGTCTGCAGCAGCTGTCCCGGCTCTGAGAGCAGCTCCAGC 1046
QY 252 AlaArgSer 254
Db 1047 GTCTCTTCG 1055

RESULT 5
US-10-029-386-3728
; Sequence 3728, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3728
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
; OTHER INFORMATION: NT HIT: g114735271, EVALUE 1.00e-123
; OTHER INFORMATION:

; OTHER INFORMATION: EST HUMAN HIT: BF969269.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P54817, EVALUE 3.70e-02
US-10-029-386-3728

Alignment Scores:
Pred. No.: 1.99e-38 Length: 509
Score: 408.50 Matches: 90
Percent Similarity: 70.55% Conservative: 13
Best Local Similarity: 61.64% Mismatches: 32
Query Match: 30.81% Indels: 11
DB: 15 Gaps: 2

US-10-078-090-151 (1-260) x US-10-029-386-3728 (1-509)

QY 3 ArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyVal 22
Db 34 AAATCGAGGGCTAGAGAGGAAAGGA-----CTCAGGGCTGG----- 72
QY 23 GlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeuArgVal 42
Db 73 -----CCTCTTCTTCCATGTGGGGCTCTCTGAGAAAAACCAAGGTCTGGGCCAG 120
QY 43 GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
Db 121 GGGAGGTGTTCCTCCACACCCCGGCACCTCCCTTCTCTTAACCTTTGTATCCCTCTCT 180
QY 63 ValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIle 82
Db 181 CCCTAGGCTCTGAAGGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGGCGCATC 240
QY 83 ProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIle 102
Db 241 CCTGGGGCTACGTCAACCAACCATCTACACCTGGGTGGACCCCGCAGGGCCGCGCATC 300
QY 103 SerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAsp 122
Db 301 TCCCCACCCCTCGGGCCTGCCCCAGCCCAACCGTGGTGGTGGTGGTGGTGGTGGTGGT 360
QY 123 ArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeuValLeu 142
Db 361 CGGAGGAGCACCCTGCACCTCTCTGCAAGAGGGGATGAGAAAAAAGGTGAGTGGGTGGGG 420
QY 143 GlyAspGlyArgSerLeu 148
Db 421 AAAGGAGGCCAGCCTCTC 438

RESULT 6
US-10-029-386-17428
; Sequence 17428, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 17428
; LENGTH: 195
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
; OTHER INFORMATION: EST HUMAN HIT: BE890168.1, EVALUE 1.00e-105
; OTHER INFORMATION: NT HIT: g114735271, EVALUE 1.00e-106
; OTHER INFORMATION: SWISSPROT HIT: P35428, EVALUE 3.70e-01
US-10-029-386-17428

Alignment Scores:		
Pred. No.:	1.15e-31	195
Score:	346.00	65
Percent Similarity:	100.00%	0
Best Local Similarity:	100.00%	0
Query Match:	26.09%	0
DB:	15	0
	Length:	
	Matches:	
	Conservative:	
	Mismatches:	
	Indels:	
	Gaps:	

US-10-078-090-151 (1-260) x US-10-029-386-17428 (1-195)

[illegible]

RESULT 7

```

US-10-177-191A-5
; Sequence 5, Application US/10177191A
; Publication No. US20030148381A1
; GENERAL INFORMATION:
; APPLICANT: Hybrigenics, Institut Pasteur
; APPLICANT: Legrain P.
; APPLICANT: Daviet L.
; APPLICANT: Petit C.
; APPLICANT: Boeda B.
; APPLICANT: EL-Amraoui A.
; TITLE OF INVENTION: More Protein-Protein Interactions In The Inner Ear
; FILE REFERENCE: B4940A
; CURRENT APPLICATION NUMBER: US/10/177,191A
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299.848
; PRIOR FILING DATE: 2001/06/21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Human
US-10-177-191A-5

```

Alignment Scores:		
Pred. No.:	8.07e-27	837
Score:	312.00	83
Percent Similarity:	51.53%	35
Best Local Similarity:	36.24%	83
Query Match:	23.53%	28
DB:	15	6
	Length:	
	Matches:	
	Conservative:	
	Mismatches:	
	Indels:	
	Gaps:	

US-10-078-090-151 (1-260) x US-10-177-191A-5 (1-837)

QY	8	GluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyValGlyIleTyrValSer	27
Db	19	GAAGGTCTCGGCCTCAGCGTGCCTGGAGGCCTGGAATTGGCTGTGGACTCTTATCTCC	78
QY	28	LeuValGluProGlySerLeuAlaGluLysGluGlyLeuArgValGlyAspGlnIleLeu	47
Db	79	CACCTCATCAAGGTGGCCAGGCAGACACGCTTGGCTTCAGGTAGGGGATGAATTGTC	138
QY	48	ArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAlaValLysAlaLeuLys	67
Db	139	CGGATCAACGGGTATCCATCTCTTCCCTGTACCCATGAGGAAGTCATCAACCTGATCCGC	198

QY	68	GlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrVal	87
DB	199	ACCAAGAAGACCGTGTCCATCAAAGTGAGACACATCGACTGATCCCTGTGAAGAGCTCT	258
QY	88	ThrAsnHisIleTyrThrTrp-----ValAspProGlnGlyArgSerIleSerProPro	105
DB	259	CCTGAGGAGTCCCTCAAATGGCAGTATGTGGAT-----CAGTTCTGTGCGAATCT	309
QY	106	SerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGluGlyAspArgArgSer	125
DB	310	GGGGGTGTG-----CGAGGTGGCTTG-----GGCTCACCTGGCAATCGGACAACC	354
QY	126	ThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsnLeuValLeuGlyAspGly	145
DB	355	-----AAGGAGAAGAAGGTGTTTATCAGTCTAGTGGGCTCT	390
QY	146	ArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThr	165
DB	391	CGGGCGCTGGGCTGCAGCATCTCCAGTGGCCCCCATCCAGAAGCCTGGCATCTTCGTGAGC	450
QY	166	GlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIleLeu	185
DB	451	CACGTGAAGCCTGGCTCCCTGTCTGCAGAGGTGGGGTTAGACACAGAGACCAGATTGTG	510
QY	186	GluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLeuLys	205
DB	511	GAAGTCAATGGCATAGACTTCACCAACCTGGACCACCAAGGAGGCTGTGAATGTCTCTGAAG	570
QY	206	SerSerArgHisLeuIleLeuThrVal-----Lys	215
DB	571	AGCAGCCGCAGCCTGACCATCTCCATCGTTGTCTGGAGCCGCCGGAGCTGTTTCATGACG	630
QY	216	AspValGlyArgLeuProHisAlaArg	224
DB	631	GACCGGGAACGGCTGGAGGAGGCACGG	657

RES. II, T. 8

```

US-10-305-720-1138
; Sequence 1138, Application US/10305720
; Publication No. US20040010136A1
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.
; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expression
; FILE REFERENCE: PA-0002-1 CON
; CURRENT APPLICATION NUMBER: US/10/305,720
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/016,434
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 1490
; SOFTWARE: PERL Program
; SEQ ID NO 1138
; LENGTH: 3071
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank ID No. US20040010136A1 GI463025
US-10-305-720-1138

```

Alignment Scores:		
Pred. No.:	2.55e-16	3071
Score:	229.00	104
Percent Similarity:	39.89%	38
Best Local Similarity:	29.21%	90
Query Match:	17.27%	124
DB:	16	19
		Gaps:
		Length:
		Matches:
		Conservative:
		Mismatches:
		Indels:

US-10-078-090-151 (1-260) X US-10-305-720-1138 (1-3071)

QY 5 LysAlaHisGluGlyLeuGlyPheSerIleArgGly-----GlySerGluHisGlyVal 22
 ::::| | | | | | | | | | | | | | | | | | | | | |
Dk 868 AAAGGGCCCTAAACGTTTAGCCTTCAGATTGCAGGAGGTGTGGGAACCAACACATTCCT 927

QY 23 Gly-----IleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
Db 928 GGAGACAACAGCATTATGTAACTAAATTTATAGTGGAGGAGCTGCACAAAAGATGGA 987
QY 40 ---LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThr 58
Db 988 AGGTTGCAAGTAGGAGATAGACTACTAATGGTAAACAACACTACAGTTTACAAGAATAACA 1047
QY 59 HisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuVal---LeuSerValTyr 77
Db 1048 CACGAAGAGGCAGTAGCAATATATAAGAACACATCAGAGGTAGTTTATTTAAAGTTGGC 1107
QY 78 SerAlaGlyArgIleProGlyTyrValThrAsnHisIleTyrThrTrpValAspPro 97
Db 1108 AACCCCACTACCATT-----TATATGACT-----GATCCT 1137
QY 98 GlnGly-----ArgSerIleSerProPro-----Ser 106
Db 1138 TATGTTCCACCTGATATTACTACTCTTTATTTCTCCACCAATGGAAACCATCTACTCTCT 1197
QY 107 Gly-----107
Db 1198 GGCAACAATGGCACTTTAGAAATATAAAACCTCCCTGCCACCCATCTCTCCAGGAGGTAC 1257
QY 108 -----LeuProGlnPro--- 111
Db 1258 TCACCAATTCCAAAGCACATGCTTGTGACGACGACTACACCAGGCCTCCGGAACTGTT 1317
QY 111 ----- 111
Db 1318 TACAGCACTGTGAACAAACTATGTGATAAGCCTGTTCTCCAGGCACACTATTCCTCTGTT 1377
QY 112 -----HisGlyGlyAlaLeu 116
Db 1378 GAGTGTGACAAAAGCTTCTCCTCTCAGCTCCCTATTCCCACTACCACTAGGCTGCTA 1437
QY 117 ArgGlnGlnGlu-----GlyAspArgArgSerThrLeu 127
Db 1438 CCTGACTCTGAGATGACCACTGATTCCTCAACATAGCACCCGCAACTCGTCAGCCTTCAATG 1497
QY 128 HisLeuLeuGlnGly-----GlyAspGluLysLysValAsnLeuValLeuGly 143
Db 1498 ACTCTCCAACGGGCGCTCTCCCTGGAGAGAGAGCCTCGCAAGGTAGTCTCTGCACAAAGGC 1557
QY 144 AspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyr 163
Db 1558 ---TCCACTGGCTGGCTTCAACATTGTCGGTGGG---GAAGATGGAGAGGTATTTT 1611
QY 164 IleThrGlyValAspProGlySerGluAlaGluGlySerGly---LeuLysValGlyAsp 182
Db 1612 GTGTCCTTCACTTGGCTGGTGGACCAAGACCTAAGTGGGGAGCTCCAGAGAGGAGAC 1671
QY 183 GlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArg 202
Db 1672 CAGATCCTATTCGGTGAATGGCATTGACCTCGGTGGTGCATCCCAAGAGCAGGAGCTGCT 1731
QY 203 LeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeuProHis 222
Db 1732 GCATAAAGGGGCTGGACAGACAGTACAGGATTATAGCACAATATCAACCTGAAGATTAC 1791
QY 223 AlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMet 242
Db 1792 GCTCGATT-----GAGGCCAAA-----ATCCATGACCTACGAGCAGATG 1833
QY 243 AlaAsn-----SerAlaGlySerGlyHisSerAlaArgSerAsn 255
Db 1834 ATGAACCAACAGCATGAGCTCGGGTCCGGA---TCCCTGCGAACCATT 1878

RESULT 9
US-10-451-207-14
; Sequence 14, Application US/10451207
; Publication No. US20040038267A1
; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: BURFORD, Neil
; APPLICANT: DING, Li
; APPLICANT: YUE, Henry
; APPLICANT: THORNTON, Michael B.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: GANDHI, Ameen R.
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: WARREN, Bridget A.
; APPLICANT: TANG, Y. Tom
; APPLICANT: KHAN, Farrah A.
; APPLICANT: YAO, Monique G.
; APPLICANT: EMERLING, Brooke M.
; TITLE OF INVENTION: INTRACELLULAR SIGNALING MOLECULES
; FILE REFERENCE: PF-0868 USN
; CURRENT APPLICATION NUMBER: US/10/451,207
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: PCT/US01/50315
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,804
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/260,102
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PERL Program
; SEQ ID NO 14
; LENGTH: 5139
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040038267A1 6942051CB1
US-10-451-207-14

Alignment Scores:
Pred. No.: 1,01e-15 Length: 5139
Score: 226.50 Matches: 82
Percent Similarity: 42.96% Conservative: 37
Best Local Similarity: 29.60% Mismatches: 81
Query Match: 17.08% Indels: 77
DB: 13 Gaps: 10

US-10-078-090-151 (1-260) x US-10-451-207-14 (1-5139)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlySerGluHis 20
Db 3013 ATCCGTCTGCCAAGAGCTGGGGCCCTCTGGGCTTAGTATTGTCGAGGCTCCGACCAT 3072
QY 21 -----GlyVal-----GlyIleTyrValSerLeuValGluProGly 32
Db 3073 TCCAGCCACCCGTTGGTGTCCAGGAGCCTGGTGTTCATCTCCAAGTGTCTCCCGCG 3132
QY 33 SerLeuAlaGluLysGluGlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLys 52
Db 3133 GGCCTGGCCGCTCGCAGCGGCTGCGGGTGGGACCGCATCTCGCAGTGAACGGGCAA 3192
QY 53 SerLeuAlaArgValThrHisAlaGluAlaValLysAla---LeuLysGlySerLysLys 71
Db 3193 GACGTGGCGGATGCCACGCCACCAAGACGACGTAGTCCCTGCTCCGGCCCTGCTGGAG 3252
QY 72 LeuValLeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsnHisIle 91
Db 3253 CTGTCGCTGCTGGTG----- 3267
QY 92 TyrThrTrpValAspProGlnGlyArgSerIleSerProSerGlyLeuProGlnPro 111
Db 3268 -----CGGAGGAGCCCGGACCCCGGGGCTTA----- 3294

QY 112 HisGlyGlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGln 131
Db 3294 ----- 3294
QY 132 GlyGlyAspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThr 151
Db 3295 -----CGGGAAGTGTGATCCAGAGGACCTGGGAGAGGCTGGGCATCAGC 3342
QY 152 IleArgGlyGlyAlaGlu-----TyrGlyLeuGlyIle 162
Db 3343 ATCCGCGGGGTGCCAGGGGCCACGCTGGCAACCCCGCGACCCACAGAGGCGATC 3402
QY 163 TyrIleThrGlyValAspPro---GlySerGluAlaGluGlySerGlyLeuLysValGly 181
Db 3403 TTCAATCTCCAGGTAGCCCCCAGCGGGGAGCGCGGCGACGGTGGCTGGTGGGT 3462
QY 182 AspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaVal 201
Db 3463 TTGGCGGTGTGGAGGTGAACCCAGCAGAGCTGCTGGGCTGACGCACGCGAGGCGGTG 3522
QY 202 ArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeuPro 221
Db 3523 CAGCTGCTCCGAGGTGGGGCGACACCTCACCGTGTGGTCTGTGACGGC----- 3573
QY 222 HisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThr 241
Db 3574 ---TTCGAGGCCAGCACCGACGCGCCCTGGAGGTGTCCCCAGGTGTCTGCAACCCC 3630
QY 242 MetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeuGlnThr 258
Db 3631 TTTGCG-----GCAGGCATCGGCCAC-----CGAACAGCCTGGAGAGC 3669
RESULT 10
US-09-919-497-9
; Sequence 9, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 3995
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-497-9
Alignment Scores:
Pred. No.: 2,42e-15 Length: 3995
Score: 222.00 Matches: 91
Percent Similarity: 46.45% Conservative: 40
Best Local Similarity: 32.27% Mismatches: 110
Query Match: 16.74% Indels: 41
DB: 9 Gaps: 13
US-10-078-090-151 (1-260) x US-09-919-497-9 (1-3995)
QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGly-----GlySer 18
Db 1469 ATCAAGCTCATCAAGGGCCCTAAAGTCTTGCTTCAGCATCCGAGGGGCGTAGGGAAC 1528
QY 19 GluHisGlyValGly-----IleTyrValSerLeuValGluProGlySerLeuAla 35
Db 1529 CAGCACATCCAGGAGATATAGCATCTATGTAAACAAGATCATCGAAGGGGTGTGTGC 1588
QY 36 GluLysGluGly---LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeu 54
Db 1589 CACAAGGATGGAGGTTGCAGATTGGAGACAAGATCTCTGGCGGTCAACAGTGTGGGGCTA 1648

QY 55 AlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLeuValLeu 74
Db 1649 GAGGACGTGATGATGAAGATGCTGTGGCAGCCCTGAAGAACACAGTATGTTGTCTAC 1708
QY 75 -----SerValTyrSerAlaGlyArgIleProGlyGlyTyr 86
Db 1709 CTAAAGGTGGCAAGCCAGCAATGCCTACCTGAGTGACAGTATGCTCCCCAGACATC 1768
QY 87 ValThrAsnHisIleTyr-ThrTrpValAspProGlnGlyArgSerIleSerProse 106
Db 1769 ACAACCTCTTATTTCCAGCACCTGGACA-----ATGAGATCAGTCACAGCAGCTAC 1819
QY 106 rGlyLeuProGln-ProHis-----GlyGlyAlaLeuArgGlnG 119
Db 1820 CTGGGCACCGACTACCCACAGCCATGACCCCACTTCCCTCGGGCGCTACTCTCCAGTG 1879
QY 119 lnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValA 139
Db 1880 GCCAAGGACCTGCTCGGGAGGAAGACAT-----TCCCCGAGAACCCGAGGCGA 1927
QY 139 sn-LeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyr 158
Db 1928 ATTGTGATCCACCCGGGCTCCACGGGCTTCAACATCGTGGGTGGC---GAGGAC 1984
QY 159 GlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly--- 177
Db 1985 GGTGAAGGCATCTTTCATCTCTTTATCTCTGGCCGGGGCCCTGCAGACCTCAGTGGGAG 2044
QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
Db 2045 CTGCGGAAGGGGACCATCTGTCTGGTCAACGCTGTGGACCTCCGAAATGCCAGCCAT 2104
QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspVal 217
Db 2105 GACGAGGCTGCCATTGCCCTGAAGAATGCGGGTCACACGGTCACGATCATCGCTCAGTAT 2164
QY 218 GlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArg 237
Db 2165 AAACAGAGAAGATACAGCCGATTC-----GAGGCCAAG-----ATCCACGAC 2206
QY 238 IleArgGluThrMetAlaAsnSerAla---GlySerGly---HisSerAlaArgSerAsn 255
Db 2207 CTTCCGGAACAGCTCATGTAACAGCAGCAGCTGGGCTCAGGGACTGCGTCTTGGCGAGCAAC 2266
RESULT 11
US-10-170-385-300
; Sequence 300, Application US/10170385
; Publication No. US2003020372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 532682000100
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300
; LENGTH: 3995
; TYPE: DNA

QY 193 LeuAsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeu 212
Db 2705 CAGAAC-----CGTGAAGAGGCTGTGGCTCTTCTTAACCAAGTGAAGAAATAAAACTTT 2652
QY 213 ThrValLysAspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrp 232
Db 2651 TCATTG-----CTGATTGCAAGGGCTGAACCTCCAGCTGGATGAG---GGCTGG 2607
QY 233 IleAlaSerSerArg 237
Db 2606 ATGGATGATGACAGG 2592
RESULT 13
US-10-296-115-496/c
; Sequence 496, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 496
; LENGTH: 4235
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-496
Alignment Scores:
Pred. No.: 4.42e-13 Length: 4235
Score: 203.00 Matches: 75
Percent Similarity: 46.42% Conservative: 48
Best Local Similarity: 28.30% Mismatches: 98
Query Match: 15.31% Indels: 44
DB: 13 Gaps: 12
US-10-078-090-151 (1-260) x US-10-296-115-496 (1-4235)
QY 10 LeuGlyPheSerIleArgGlyGly-----SerGluHisGly 21
Db 3365 CTGGGATTCAATATTATTGGTGGCGCGGAGTGTGGATAACCAAGATGGATCATCCAGT 3306
QY 22 ValGlyIleTyrValSer---LeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 3305 GAAGGAATCTTTGTATCCAAAGATAGTTGACAGTGGGCTGCAGCCCAAGGAAGAGGCCTG 3246
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 3245 CAATTCATGACAGGATTATTGAGGTCAACGGCAGAGACTTATCCAGAGCAACTCATGAC 3186
QY 61 GluAlaValLysAlaLeuLysGlySerLysLys---LeuValLeuSerValTyr----- 77
Db 3185 CAGGCTGTGGAAGCTTTCAGACAGACAGCCCAAGGAGCCCATAGTGGTGCAGGTGTTGAGAAGA 3126
QY 78 -----SerAlaGlyArgIleProGlyGlyTyr 86
Db 3125 ACACCAAGGACCAAAATGTTACGCGCTCCATCAGAGTCTCAGCTGGTGGACACGGGAACC 3066
QY 87 ValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIleSerPro----- 104
Db 3065 CAACCGACATCACCTTTGAACATATCATGCGCCCTCACTAAGATGTCTCTCCAGCCCA 3006
QY 105 -----ProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlu 120
Db 3005 CCCGTGCTGGATCCCTATCTCTCCAGAGGAGCATCCCTCAGCCCATGAATACTACGAT 2946
QY 121 ---GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsn 139

Db 2945 CCAAATGACTACATTGGAGACATCCATCAGAGATGGACAGGAGGAGCTGGAGCTGGAG 2886
QY 140 LeuVal-----LeuGlyAspGlyArgSerLeuGlyLeuThrIle-----Arg 153
Db 2885 GAAGTGGACCTCTACAGAATGAACAGCCAGGACAAGCTGGCCCTCACTGTGTGTACCGG 2826
QY 154 GlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAla 173
Db 2825 ACGGACGATGAAGACGACATGGGATTTATATCATGATGAGATTGACCCCTAACAGCATTGCA 2766
QY 174 GluGlySerGly---LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPhe 192
Db 2765 GCCAAGGATGGCGCATCCGAGAAGGAGAGCCGCAATATCCAGATTAATGGGATAGAGGTG 2706
QY 193 LeuAsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeu 212
Db 2705 CAGAAC-----CGTGAAGAGGCTGTGGCTCTTCTTAACCAAGTGAAGAAATAAAACTTT 2652
QY 213 ThrValLysAspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrp 232
Db 2651 TCATTG-----CTGATTGCAAGGGCTGAACCTCCAGCTGGATGAG---GGCTGG 2607
QY 233 IleAlaSerSerArg 237
Db 2606 ATGGATGATGACAGG 2592
RESULT 14
US-09-964-956-24
; Sequence 24, Application US/09964956
; Publication No. US20040043926A1
; GENERAL INFORMATION:
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shinkets, Richard A
; TITLE OF INVENTION: No. US20040043926A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/09/964,956
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,631
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,633
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,808
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,064
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,065
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,066
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,135
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,434
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,321
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/238,399
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/238,396
; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/276,667
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/294,823
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/304,868
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 8640
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-956-24

Alignment Scores:
Pred. No.: 4.51e-12 Length: 8640
Score: 198.00 Matches: 87
Percent Similarity: 36.72% Conservatives: 43
Best Local Similarity: 24.58% Mismatches: 124
Query Match: 14.93% Indels: 100
DB: 13 Gaps: 13

US-10-078-090-151 (1-260) x US-09-964-956-24 (1-8640)

QY 5 LysAlaHisGluGlyLeuGlyPheSerIleArgGlyGly-----SerGlu 19
DB 1121 AAAGAATCGGATGGGCTGGGATTCAGGTTAGTGAGGCGGAGGATCAAGCGCTCACCT 1180
QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGlu-----Lys 37
DB 1181 CAC-----GCTATCGTTGTCACTCAAGTGAAGGAAGGAGGTCCGCTCACAGGCTCAGG 1234
QY 38 GluGly---LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArg 56
DB 1235 GATGGCAGCGTGTCTTAGGAGATGAGCTGTGTTAATCAATGTTCACTTACTGTCGGG 1294
QY 57 ValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuVal-----73
DB 1295 CTCTCCACGAGGAAGCAGTGGCCATTCTTCGCTCCGCCCGGAATGGTGCAGCTGTG 1354
QY 74 -----LeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsn 89
DB 1355 GTGGCCAGCAAGGTAGGTGTCTTCTTCATTTTCAGATGCTCGGACAGATGAACCCCAA 1414
QY 90 HisIleTyrThrTrpValAspProGlnGlyArgSerIleSerProProSerGly-----107
DB 1415 GATGTGTGCGGTGCTGAGGAATCCAAAGGGAACCTTGAAAGTCCCAACAGGGCAGCAAT 1474
QY 108 -----LeuProGlnProHis 112
DB 1475 AAAATCAAGCTCAAGAGTCGCCTTTACGGTAGGTGGGGCTCTACCTGATGCAGCCTGTC 1534
QY 113 GlyGlyAlaLeuArg-----117
DB 1535 GGGGGTGTACACCGCCTTGAGTCAGTTGAAGAATATAACAGAGTCATGGTGGGAATGGG 1594
QY 118 -----GlnGlnGluGlyAspArgArgSerThrLeuHis 128
DB 1595 GACCCCGGATCCGGATGTTGGAGGTCTCCGAGATGCCGGAACACTCCCTCCCGCAG 1654
QY 129 LeuLeuGlnGlyAspGluLysLys-----137
DB 1655 CTGCTGGACTCTTCAGTGCCTCACAGGAATACCACATTGTGAAGAAGTCTACCGCTCC 1714
QY 137 -----137
DB 1715 TTAAGCAGCACTCAGGTGGAATCTCCTTGAGGCTCATTCGGCCATCCGTCATCTCGATC 1774
QY 136 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
DB 1775 ATTGGTTGTACAAAGAAAAGCAAGGCCCTTGCTTTAGTATTGTCGAGGTCGAGAC 1834
QY 158 Tyr-----GlyLeuGlyIleTyrIleThrGlyValAspPro---GlySerGluAla 173

DB 1835 TGCATTCTGGACAGATGGGGATTTTGTCAAGACCATCTTCCAAATGGATCAGCTGCA 1894
QY 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
DB 1895 GAGGACGGAAGACTTAAAGAAGGTGATGAATCCTAGATGTAATGAATACCAATAAAG 1954
QY 194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHis-----LeuIle 211
DB 1955 GGCTTGACATTTCAAGAAGCCATTTCATACCTTTTAAGCAAAATCCGGAGTGGATTATTGTT 2014
QY 212 LeuThrValLysAspValGlyArgLeuProHisAlaArg-----ThrThrValAspGlu 229
DB 2015 TTAACGGTACGCACAAAGTTGGTGAGCCCCCAGCCTCACACCTGCTCGACACCCACAC 2074
QY 230 ThrLysTrpIleAlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGly-----247
DB 2075 ATGACGAGATCCGCCTCCCGAACTTCAATACCAATGCGGGAGCCTCGGCGGAGGTTCC 2134
QY 248 ---SerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
DB 2135 GATGAAGGCAGTTCTTCATCCTCCCTGGGTGCGGAAGACCCCTGGG 2176

RESULT 15
US-09-964-956-26
; Sequence 26, Application US/09964956
; Publication No. US20040043926A1
; GENERAL INFORMATION:
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shimkets, Richard A
; TITLE OF INVENTION: No. US20040043926A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/09/964,956
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,631
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,633
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,808
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,064
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,065
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,066
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,135
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,434
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,321
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/238,399
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/238,396
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/276,667
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/294,823

```

; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/304,868
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 8640
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-956-26

Alignment Scores:
Pred. No.:      4.51e-12      Length:      8640
Score:          198.00      Matches:      87
Percent Similarity: 36.72%      Conservative: 43
Best Local Similarity: 24.58%      Mismatches: 124
Query Match:      14.93%      Indels:      100
DB:              13          Gaps:       13

US-10-078-090-151 (1-260) x US-09-964-956-26 (1-8640)

QY      5 LysAlaHisGluGlyLeuGlyPheSerIleArgGlyGly-----SerGlu 19
Db      1121 AAAGAATCGGATGGCTGGGATTCAGGTTAGTGAGGCCGAGGATCAAAAGCGCTCACCT 1180

QY      20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGlu-----Lys 37
Db      1181 CAC-----GCTATCGTTGTCTCACTCAAGTGAAGGAAGGAGGTGCCGCTCACAGGCTCAGG 1234

QY      38 GluGly---LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArg 56
Db      1235 GATGGCAGGCTGTCTTAGGAGATGAGCTGTGGTAATCAATGGTCATTTACTGGTCGGG 1294

QY      57 ValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuVal-----Lys 73
Db      1295 CTCTCCACGAGGAAGCAGTGGCCATTCTTCGCTCCGCCACGGGAATGGTGAGCTGTG 1354

QY      74 -----LeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsn 89
Db      1355 GTGGCCAGCAAGGTAGGTGTCTTCTGCTATTTTCAGATGCTGGACAGATGAACCCCAA 1414

QY      90 HisIleTyrThrTrpValAspProGlnGlyArgSerIleSerProProSerGly----- 107
Db      1415 GATGTGTGGGTGCTGAGGAATCCAAAGGGGAACCTTGGAAAGTCCCAAACAGGGCAGCAAT 1474

QY      108 -----LeuProGlnProHis 112
Db      1475 AAAATCAAGCTCAAGAGTCGCCCTTTACGGTAGGTGGGGCTCTACCTGATGCAGCCTGTC 1534

QY      113 GlyGlyAlaLeuArg----- 117
Db      1535 GGGGGTGTTACACCGCCTTGAGTCAGTTGAAGAATATAACGAGTGATGGTGCGGAATGGG 1594

QY      118 -----GlnGlnGluGlyAspArgArgSerThrLeuHis 128
Db      1595 GACCCCGGATCCGGATGTTGGAGGTCTCCCGAGATGGCCGGAACACTCCCTCCCGCAG 1654

QY      129 LeuLeuGlnGlyAspGluLysLys----- 137
Db      1655 CTGCTGGACTCTCCAGTGCCTCACAGGAATACCACATTGTGAAGAAGTCTACCCGCTCC 1714

QY      137 ----- 137
Db      1715 TTAAGCAGACTCAGGTGAATCTCCTCGAGGCTCATTCGGCCATCCGTCATCTCGATC 1774

QY      138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
Db      1775 ATTGGGTTGTACAAAGAAAAAGGCAAGGCCCTTGGCTTTAGTATTGCTGAGGTCGAGAC 1834

QY      158 Tyr-----GlyLeuGlyIleTyrIleThrGlyValAspPro---GlySerGluAla 173
Db      1835 TGCATTCTGTCGACAGATGGGGATTTTGTCAAGACCATTCTTCCCAATATGGATCAGCTGCA 1894
```

```

QY      174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
Db      1895 GAGGACGGAAGACTTAAAGAAAGGGGATGAAATCCTAGATGTAATGGAATACCAATAAAG 1954

QY      194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHis-----LeuIle 211
Db      1955 GGCTTGACATTTCAAGAAGCCATTTCATACCTTTAAGCAAAATCCGGAGTGGATTATTGTT 2014

QY      212 LeuThrValLysAspValGlyArgLeuProHisAlaArg-----ThrThrValAspGlu 229
Db      2015 TTAACGGTACGCACAAAGTTGGTGAGCCCCCAGCCCTCACACCCCTGCTCGACACCCACAC 2074

QY      230 ThrLysTrpIleAlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGly----- 247
Db      2075 ATGAGCAGATCCGCCTCCCCGAACTTCAATACCAGTGGGGGAGCCTCGCGGGAGGTTCC 2134

QY      248 ---SerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
Db      2135 GATGAAGGCAGTTCTTCATCCCTGGGTCCGAAGACCCCTGGG 2176
```

Search completed: April 22, 2004, 03:48:24
Job time : 407 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 22, 2004, 00:07:22 ; Search time 2513 Seconds
(without alignments)
3089.601 Million cell updates/sec

Title: US-10-078-090-151
Perfect score: 1326
Sequence: 1 LRRKAHEGLFGSIRGGSEH.....TWANSAGSGHSARSNLQTPG 260

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55036578

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2 1/USPTO spool p/US10078090/runat 20042004 132810 16601/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10078090@cgn 1 1 4237 @runat 20042004 132810 16601 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_estl.*
10: gb_est2.*
11: gb_est3.*
12: gb_est4.*
13: gb_est5.*
14: gb_estfun.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1261.5	95.1	4569	11	BC014524	BC014524 Homo sapi
2	1015	76.5	602	10	BE890168	BE890168 601513147
3	922.5	69.6	651	12	BF732824	BF732824 603353931
4	896.5	67.6	905	13	BX390092	BX390092 BX390092
5	860	64.9	729	14	CA321445	CA321445 UI-M-FW0-
6	686	51.7	693	10	BF969269	BF969269 602269794
7	662	49.9	904	14	CB209310	CB209310 AGENCOURT
8	593.5	44.8	643	14	CD310081	CD310081 StrPu691.
c	581	43.8	1160	14	CF661317	CF661317 CcLM09a34
10	564	42.5	396	14	CB773455	CB773455 AMGNNUC:T
11	553.5	41.7	870	14	CD760225	CD760225 AGENCOURT
12	526	39.7	374	14	CB692138	CB692138 AMGNNUC:T
13	523	39.4	343	14	CB693399	CB693399 AMGNNUC:S
14	519	39.1	383	10	BB872069	BB872069 BB872069
c	512	38.6	325	10	BF953476	BF953476 RC3-NN118
16	509	38.4	324	10	BF953475	BF953475 RC3-NN118
17	504	38.0	342	10	BF953545	BF953545 RC3-NN118
18	497	37.5	343	10	BF953479	BF953479 RC3-NN118
19	488	36.8	323	10	BF953542	BF953542 RC3-NN118
20	488	36.8	342	10	BF953472	BF953472 RC3-NN118
21	485	36.6	675	14	CA327501	CA327501 UI-M-FY0-
c	482	36.3	308	10	BF953480	BF953480 RC3-NN118
23	478	36.0	1174	13	BU166723	BU166723 AGENCOURT
c	460	34.7	347	10	BF953470	BF953470 RC3-NN118
25	429	32.4	713	13	BX880758	BX880758 BX880758
26	429	32.4	736	13	BX885444	BX885444 BX885444
27	429	32.4	761	13	BX868932	BX868932 BX868932
c	425	32.1	362	10	BF953546	BF953546 RC3-NN118
29	421	31.7	310	10	BF953522	BF953522 RC3-NN118
30	401	30.2	267	14	CA949304	CA949304 it44b01.Y
31	363.5	27.4	1018	12	BM452626	BM452626 AGENCOURT
c	361.5	27.3	334	10	BF953532	BF953532 RC3-NN118
33	357	26.9	799	12	BG862377	BG862377 602796268
34	355.5	26.8	534	10	AW742499	AW742499 up58a09.Y
35	349.5	26.4	586	12	BM940881	BM940881 UI-M-CG0p
36	334	25.2	531	10	BF074654	BF074654 222121.MA
37	319	24.1	641	28	AZ625547	AZ625547 1M0465B12
38	302	22.8	1139	13	BX344632	BX344632 BX344632
39	297.5	22.4	668	13	BU610146	BU610146 UI-M-DJ1-
c	285.5	21.5	1201	9	AL556078	AL556078 AL556078
41	260.5	19.6	888	29	CNS041G2	AL270155 Tetraodon
c	260.5	19.6	961	29	CNS043K2	AL272891 Tetraodon
c	260	19.6	305	9	AA938876	AA938876 op74b06.s
44	259.5	19.6	1045	29	CNS04F1L	AL287778 Tetraodon
45	234	17.6	907	13	BQ433907	BQ433907 AGENCOURT

ALIGNMENTS

RESULT 1
BC014524
LOCUS
DEFINITION
Homo sapiens CASK-interacting protein CIP98, mRNA (cDNA clone IMAGE:3834205), with apparent retained intron.
ACCESSION
BC014524
VERSION
BC014524.1 GI:17939549
KEYWORDS
HTC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 4569)

AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 4569)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 14 Row: f Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: retained intron.

FEATURES
source
1..4569
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3834205"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH_MGC_9"
/lab_host="DH10B-R"
/note="vector: pOTB7"

ORIGIN
Alignment Scores:
Pred. No.: 1.49e-101 Length: 4569
Score: 1261.50 Matches: 250
Percent Similarity: 96.17% Conservative: 1
Best Local Similarity: 95.79% Mismatches: 9
Query Match: 95.14% Indels: 1
DB: 11 Gaps: 1

US-10-078-090-151 (1-260) x BC014524 (1-4569)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
|||||

Db 1582 TTGCGGCGTGCCCAAGGCCCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGGCTCGGAGCAC 1641

QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
|||||

Db 1642 GGCCTGGGCATCTACGTGTCTCTGGTGGAACAGGCTCTCTAGCTGAGAAGGAAGGACTG 1701
|||||

QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
|||||

Db 1702 CGGGTCGGGGACCAAGATTCTGCGCTCAACGACAAATCCCTGGCCCCGGGTGACCCACGCG 1761
|||||

QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
|||||

Db 1762 GAGGCCGTCAAGGCTCTGAAGGGCTCCAAGAAGTGGTGTCTGTGTGTACTCAGCAGGG 1821
|||||

QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
|||||

Db 1822 CGCATCCCTGGGGCTACGTCAACCAACCATCATACCTGGGTGGACCCCGAGGGCCGC 1881
|||||

QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlu 120
|||||

Db 1882 AGCATCTCCCAACCTCGGGCTGCCCCAGCCCCAGCTGGTGGTGGTGGTGGTGGTGGTGG 1941
|||||

QY 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
|||||

Db 1942 GGTGACCGGAGGAGCACCCCTGCACCTCTCTCAAGGAGGGGATGAGAAAAAGGTGAACCTG 2001
|||||

QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
|||||

Db 2002 GTGCTGGGGACGGCCGCTCCCTGGGCCTCACGATCCGTGGGGAGCTGAGTACGGCCTT 2061
|||||

QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
|||||

Db 2062 GGCATTATACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAAAGGCAGCGGGCTCAAGGTT 2121
|||||

QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
|||||

Db 2122 GGGGACCAAGATTCTAGAAGTGAATGGCGGAGCTTTCTCAACATCTTACACGACGAGGCT 2181
|||||

QY 201 ValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
|||||

Db 2182 GTCAGGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTG 2241
|||||

QY 221 ProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGlu 240
|||||

Db 2242 CCCCATGCCCGACCACTGTGGACGAGACCAAGTGATCGCCAGTTCGCCGATCAGGGAG 2301
|||||

QY 241 ThrMetAlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrPro 259
|||||

Db 2302 ACCATGGCGAACTCGGCAGGGTTTCTTGGCGATCTCACACAGAGAAGGAATAACAAGCCA 2361
|||||

QY 260 Gly 260
|||||

Db 2362 GGA 2364

RESULT 2
BE890168
LOCUS 601513147F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914675 5',
DEFINITION mRNA sequence.
ACCESSION BE890168
VERSION BE890168.1 GI:10348220
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM9736 row: m column: 12
High quality sequence stop: 602.
Location/Qualifiers
1. .602
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3914675"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb."
Alignment Scores:
Pred. No.: 9.27e-81 Length: 602
Score: 1015.00 Matches: 200
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.55% Indels: 0
DB: 10 Gaps: 0

ORIGIN

US-10-078-090-151 (1-260) x BE890168 (1-602)

QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
DB 2 GCGGTGGGCATCTACGTGCTCTCTGGTGAACCAAGGCTCTCTAGCTGAGAAGGAGGACTG 61
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
DB 62 CGGGTCGGGACCAGATTCTCGCGCTCAACGACAAATCCCTGGCCCGGCTGACCCACGCG 121
QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
DB 122 GAGGCCGTCAAGGCTCTGAAGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCAGGG 181
QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
DB 182 CGCATCCCTGGGGCTACGTCAACCAACCATCTACACCTGGTGGACCCCGCAGGGCCGC 241
QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
DB 242 AGCATCTCCCCACCCTCGGGCCTGCCCCAGCCCCACGGTGTGGCTGAGGCAGCAGGAG 301
QY 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
DB 302 GGTGACCGGAGGAGCACCCCTGCACCTCCTGCAAGGAGGGGATGAGAAAGGTGAACCTG 361
QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
DB 362 GTGCTGGGGACCGGCCGTCCCTGGGCCTCAGCATCCGTGGGGGAGCTGAGTACGGCCTT 421
QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
DB 422 GGCATTATACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAAAGGACGGGGCTCAAGTT 481
QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
DB 482 GGGGACCAGATTCTAGAAGTGAATGGCGGAGCTTCTCAACATCCTACACGACGAGGCT 541
QY 201 ValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
DB 542 GTCAGGCTGCTTAAGTCATCTCGGCACCTCATCCTGACAGTGAAGGACGTCGGGAGGCTG 601

RESULT 3

BI732824
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

BI732824 651 bp mRNA linear EST 20-SEP-2001
603353931F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5361582 5',
mRNA sequence.
BI732824
BI732824.1 GI:15709837
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 651)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM11919 row: m column: 07
High quality sequence stop: 651.
Location/Qualifiers
1. .651
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5361582"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_94"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 1.96e-72 Length: 651
Score: 922.50 Matches: 187
Percent Similarity: 90.57% Conservative: 5
Best Local Similarity: 88.21% Mismatches: 19
Query Match: 69.57% Indels: 2
DB: 12 Gaps: 1

US-10-078-090-151 (1-260) x BI732824 (1-651)

QY 50 AsnAspLysSerLeuAlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySer 69
DB 3 AACGATAAATCTCTAGCCCGGTTGACCCACGCGAGGCTGTCAAGGCTCTCAAAGGCTCC 62
QY 70 LysLysLeuValLeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsn 89
DB 63 AAGAAGCTGGTGTCTGTATATACTCAGCTGGGCGTATCCCAGG-GGCTATGTGACCAAC 121
QY 90 HisIleTyrThrTrpValAspProGlnGlyArgSerIleSerProProSerGlyLeuPro 109
DB 122 CACATCTACACCTGGGTGGACCCACAGGGTTCGAAGCACATCCCCCTCCCTCCAGCCTGCC 181
QY 110 GlnProHisGlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeu 129
DB 182 CAGCCCCATGGCAGCACCTTGAGACAGCGTGAAGATGACCGAAGGAGTACCCCTCCACCTC 241
QY 130 LeuGlnGlyGlyAspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGly 149
DB 242 CTGCAGAGTGGAGATGAGAAAAAGGTGAACCTGGTGTGGGGACGGCCGGTCTCTGGGC 301
QY 150 LeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspPro 169

Db 302 CTCACGATCCGAGGTGGAGCAGAGTACGGCCTTGGCATTTACATCACTGGTGTGGACCCA 361
Qy 170 GlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGly 189
Db 362 GGCTCTGAAGCAGAGAGCAGCGGCCTCAAGGTTGGAGACCAGATTCTGGAGTGAATGGG 421
Qy 190 ArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLysSerSerArgHis 209
Db 422 CCGAGCTTCTCAACATCCTGCATGATGAGGCAGTGAAGCTGCTCAAGTCATCCCGGCAC 481
Qy 210 LeuIleLeuThrValLysAspValGlyArgLeuProHisAlaArgThrThrValAspGlu 229
Db 482 CTCATCCTGACGGTGAAGGACGTCGGAAGGCTGCCCCACGCACGTACCAACCGTGGACCAG 541
Qy 230 ThrLysTrpIleAlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGly---Ser 248
Db 542 ACCAAGTGGATCGCCAGTTCCTCGGATCGGGAAAGCGTCGCCCAACTCAGCAGGGTTTCCA 601
Qy 249 GlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
Db 602 GGGGACCACACAGAAGAAGGACAAAGCAAGCCAGGA 637

RESULT 4
BX390092 905 bp mRNA linear EST 08-MAY-2003
LOCUS BX390092 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DK007YH03 5-PRIME, mRNA sequence.
ACCESSION BX390092
VERSION BX390092.1 GI:30463208
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 905)
TITLE Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5156.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAF042ZH04_AF03969_1&cluster=5156.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0BAF042ZH04_AF03969_1.

FEATURES
Source Location/Qualifiers
1..905
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK007YH03"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores: 6.67e-70 Length: 905
Pred. No.: 896.50 Matches: 186
Score: 95.00% Conservative: 4
Best Local Similarity: 93.00% Mismatches: 6
Query Match: 67.61% Indels: 5
DB: 13 Gaps: 2

US-10-078-090-151 (1-260) x BX390092 (1-905)
Qy 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 313 TTGCGGCGTGCCCAAGGCCACAGAGGGCTTGGGCTTCAGCATCCGTGGGGGCTCGGAGCAC 372
Qy 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 373 GGCCTGGGCATCTACGTGTCTCTGGTGGAAACCAAGCTCTCTAGCTGAGAAAGGAGACTG 432
Qy 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 433 CGGTCGGGGACCAAGATTCTGCGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACGGC 492
Qy 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 493 GAGGCGGTCAAGGCTCTGAAGGGCTCCAAGAAGCTGGTGTGTGTGTACTACTACGAGGG 552
Qy 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 553 CGCATCCCTGGGGGTACGTACCAACCAACCATCTACACCTGGGTGGACCCCGCAGGCCGC 612
Qy 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 613 AGCATCTCCCAACCCCTCGGGCTGCCCCAGCCCCACGGTGGTGCCTGAGGCAGCAGGAG 672
Qy 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
Db 673 GGTGACCGGAGGAGCACCCCTGCACCTNCTGCA-GGAGGGGATGAGAAAAGGTGAACCTG 731
Qy 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
Db 732 GTGCTGGGGACGGCCGCTCCCTGNGCCTCACGATCCGTGGGGAGCTGAGTACGGCCTT 791
Qy 161 GlyIleTyr-IleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVa 180
Db 792 GGCATTTACATNCACTGGCGTGGACCCAGGCTCTGAAGCAGAAAGCAGCGGGCTCAGG 851
Qy 180 IGlyAspGlnIleLeuGlu---ValAsnGlyArgSerPheLeuAsnIleLeuHisAsp 198
Db 852 TGGNGACCAGATTCTAGAGTGAATGGCGGGCTTTCT-----CACATCTACACGAG 903
RESULT 5
CA321445 729 bp mRNA linear EST 09-JUL-2003
LOCUS UI-M-FW0-cbz-b-04-0-UI.r1 NIH BMAP_FW0 Mus musculus cDNA clone
DEFINITION IMAGE:6816389 5', mRNA sequence.
CA321445
ACCESSION CA321445.1 GI:24539543
VERSION EST.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 729)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
The following repetitive elements were found in this cDNA sequence: 146-183, >GC_rich#Low_complexity
Seq primer: pyX-5.

FEATURES
source
Location/Qualifiers
1. .729
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5816389"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP FWO"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Alignment Scores:
Pred. No.: 9.03e-67 Length: 729
Score: 860.00 Matches: 168
Percent Similarity: 96.02% Conservative: 1
Best Local Similarity: 95.45% Mismatches: 7
Query Match: 64.86% Indels: 0
DB: 14 Gaps: 0

US-10-078-090-151 (1-260) x CA321445 (1-729)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 201 CTGCGGCGGCCAAGGCCACGAGGGCTTGGCTTCAGCATCCGCGGGGCTCGGAACAC 260

QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 261 GGC GTGGGCATCTACGTGTCTCTAGTGGAGCCGGCTCCCTGGCAGAGAGGAGGGTTG 320

QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 321 CGGGTCGGGACCAGATTCTTGGCGGTCAACGATAAATCTTAGCCCCGGGTGACCCACGGC 380

QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 381 GAGGCTGTCAAGGCTCTCAAAGGCTCCAAAGAGCTGGTGTCTGTATACTCAGCTGGG 440

QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 441 CGTATCCAGGGGGCTATGTGACCAACACATCTACACCTGGGTGGACCCACAGGTCGA 500

QY 101 SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 501 AGCATCCCTCCCTCCAGCCTGCCCCAGCCCCCATGGCAGACCCCTGAGACAGCGTGAA 560

QY 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsnLeu 140
Db 561 GATGACCGAAGAGTAGTACCCTCCACCTCCTGCAGAGTGGAGATGAGAAAAAGGTGAACCTG 620

QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
Db 621 GTGTTGGGGACGGCCGGTCTCTGGGCTCACGATCCGAGGTGGAGCAGAGTACGGGCTT 680

QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySer 176
Db 681 GGCATTTACATCACTGGTGTGGACCCAGGCTCTGAAGCAGAGAGACGAGC 728

RESULT 6
BF969269
LOCUS
DEFINITION
602269794F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4358158 5',
mRNA sequence.
ACCESSION
BF969269
VERSION
BF969269.1 GI:12336484
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 693)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM995 row: k column: 23
High quality sequence stop: 643.
Location/Qualifiers
1. .693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4358158"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_84"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 3.11e-51 Length: 693
Score: 686.00 Matches: 140
Percent Similarity: 95.95% Conservative: 2
Best Local Similarity: 94.59% Mismatches: 6
Query Match: 51.73% Indels: 1
DB: 10 Gaps: 0

US-10-078-090-151 (1-260) x BF969269 (1-693)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 36 TTGCGGCGTGCCAAAGGCCACGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 95

QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 96 GGC GTGGGCATCTACGTGTCTCTGTGTGGAACCAAGGCTCTCTAGCTGAGAAGGAGGACTG 155

QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 156 CGGGTCGGGACCAGATTCTTGAAGGGCTCCAAAGAGCTCCCTGGCCCGGGGTGACCCACGGC 215

QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 216 GAGGCGGTCAAGGCTCTGAAGGGCTCCAAAGAGCTGGTGTGTGTGTACTCAGCAGGG 275

QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 276 CGCATCCCTGGGGGCTACGTACCAACACACATCTACACCTGGGTGGACCCGCA-GGCCCGC 334


```

/db_xref="taxon:7668"
/clone="MPMGp691D1338;MPI_SURUDI_38D13"
/tissue_type="whole larva"
/dev_stage="larva 2-3 weeks"
/lab_host="E.coli, XL1 blue"
/clone_lib="Sea urchin larva cDNA library MPMGp691"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI; Random
primed and directionally cloned in pSport1 vector using a
NotI (5'-pGACTAGTTCTAGATCGGAGCGCGCCC (T)15-3' and a
SalI 5'-TCGACCCACGCGTCCG-3'adapters (Gibco BRL)"

ORIGIN
Alignment Scores:
Pred. No.: 5.29e-43 Length: 643
Score: 593.50 Matches: 119
Percent Similarity: 75.00% Conservative: 40
Best Local Similarity: 56.13% Mismatches: 50
Query Match: 44.76% Indels: 3
DB: 14 Gaps: 1

US-10-078-090-151 (1-260) x CD310081 (1-643)

QY 16 GlyGlySerGluHisGlyValGlyIleTyrValSerLeuValGluProGlySerLeu--A 35
Db 14 GGTGGGCGGAGCATTCGGTGGGATCTTTGTGTCGTTGGTGAACCAACAGTTATGGG 73

QY 35 laGluLysGluGlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuA 55
Db 74 CAGATAAAAGAGGGTTAATCAAAAGGAGACCAGATATATGACAGTCAATGACATTCCATTG 133

QY 55 laArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeus 75
Db 134 AGAAGGTAGCCACTCTGATGCAGTGAAGATCTTAAAGCGGTCAACACAGTGTCTTGT 193

QY 75 erValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpV 95
Db 194 ATGTCAAGAGTGTGGGCGTGTCCCTGGTCCCTTCTTCTCCACAGACCTACACATGGG 253

QY 95 alAspProGlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGlyA 115
Db 254 TGAACCCCAAGGGCGCAGTGTGCCCTCCACCGACGTC---GACCCCGCTTGGTGGGA 310

QY 115 laLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspG 135
Db 311 GGATGCTAAATGATACACAGAATAGAAAGAGTGGTCTCAACCTCCTCAAGAACGGCGATG 370

QY 135 luLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyG 155
Db 371 AAAAGAAGGTCAATGTTGTAGTGAATGAAGGAGAGAGTCTTGGTTTGTATGATTCGAGGAG 430

QY 155 lyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluG 175
Db 431 GAAAGGAGTTTGGTCTTGGTATCTTCATCACTGGAATTGATACTTACTCTGTGGCAGACC 490

QY 175 lySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnI 195
Db 491 ATGCCAGTCTTAAGGTTGGCGATCAAAATCCTCGATGTCAACTCCAGGAACCTCCTCGACA 550

QY 195 leLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrVall 215
Db 551 TCGAGCACCAAGACGAGTGACATCCTCAAATCATCAAGTGAATGATGATGATGACCATCA 610

QY 215 ysAspValGlyArgLeuProHisAlaArgThr 225
Db 611 AGGACGTTGGCAAGTTGCCGTACGGCCGGACG 642

RESULT 9
CF661317/c 1160 bp mRNA linear EST 07-OCT-2003
LOCUS CcLM09a34k21f1 Carp muscle library 1 Cyprinus carpio cDNA clone
DEFINITION 34x21 5', mRNA sequence.
ACCESSION CF661317
VERSION CF661317.1 GI:37558475
```

```

KEYWORDS EST.
SOURCE Cyprinus carpio (common carp)
ORGANISM Cyprinus carpio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Cyprinus.
REFERENCE 1 (bases 1 to 1160)
AUTHORS Gracey,A.Y., Fraser,E., Li,W. and Cossins,A.R.
TITLE Microarray and EST analysis of the carp (Cyprinus carpio)
transcriptome during environmental stress
JOURNAL Unpublished (2003)
COMMENT Contact: Andrew R. Cossins
Laboratory for Environmental Gene Regulation
University of Liverpool
School of Biological Sciences, The Biosciences Building, Crown
Street, Liverpool, United Kingdom, L69 7ZB
Tel: +44(0)151-795-4510
Fax: +44(0)151-795-4431
Email: cossins@liv.ac.uk
Vector has been trimmed from this EST.
Plate: 34 row: k column: 21
Seq primer: Triplex 5' LD (5'-CTCGGAAGCGCCATTGTGTGGT-3')
High quality sequence start: 35
High quality sequence stop: 917.
FEATURES
Location/Qualifiers
1..1160
/organism="Cyprinus carpio"
/mol_type="mRNA"
/db_xref="taxon:7962"
/clone="34k21"
/sex="Male & female"
/tissue_type="Muscle"
/dev_stage="Adult"
/lab_host="E.coli Electromax DH10B"
/clone_lib="Carp muscle library 1"
/note="Vector: pTriplex2; Site 1: SfiI GGCCATTACGGCC;
Site_2: SfiI GGCGCCTCGGCC; Serially substracted cDNA
library prepared from muscle of warm, cold and hypoxia
challenged animals"

ORIGIN
Alignment Scores:
Pred. No.: 1.62e-41 Length: 1160
Score: 581.00 Matches: 120
Percent Similarity: 71.86% Conservative: 23
Best Local Similarity: 60.30% Mismatches: 41
Query Match: 43.82% Indels: 15
DB: 14 Gaps: 2

US-10-078-090-151 (1-260) x CF661317 (1-1160)

QY 70 LysLysLeuValLeuSerValTyrSerAlaGlyArg-----IleProGly 84
Db 1024 AAAAGGCTTTTAAAGAGGCTCTTTAAGTCCGGTCGGTTCAATGGGACGGATTCCAGGT 965

QY 85 GlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIleSerPro 104
Db 964 GGTACGTGACCAATCAGGGTTACACATGGGTGGACCCCTCAGGGCCGAGTGTGCCCA 905

QY 105 ProSerGlyLeuProGlnProHis-GlyGlyAlaLeuArgGlnGlnGluGlyAspArgAr 124
Db 904 CCACCAGACTTGTGGAGCAGCATGGGCTCAGGAGGAAGACATTTGGAAATCCAGCAACT 845

QY 124 gSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsnLeuValLeuGlyAs 144
Db 844 TGGCCACACCCACATTCGGCACCGAGGGTGTGAGAGAAGGTGAATATCTCACTGGATCA 785

QY 144 pGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIl 164
Db 784 TGGGCGATCTCTGGGTTTGTATCATCCGTGGAGGTGCTGAGTATGCTCTGGGCATTTATAT 725

QY 164 eThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIl 184
```

Db 724 CACCGGAGTGGACCGAGGGTCTGCTGCAGAGTACAGCGGACTAAAGGTAGGGGATCAGAT 665

QY 184 eLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLe 204

Db 664 TCTGGAGTCAATGACGATGCTCCGAGCATTTCTCATGACGAGGCCGTGCAGATCTT 605

QY 204 uLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeuProHisAlaAr 224

Db 604 GAAGAATTCCCGGCACATGCTGATGACCATAAAGGATGTGGGCGGCTGCCTCAGCACG 545

QY 224 gThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMetAlaAs 244

Db 544 AACTGTGTGGATGAGACCAATGGATTCCAGCGGCACAGATCGCTGAGAGCTCGGCCAA 485

QY 244 nSer-----AlaGlySerGlyHisSerAlaArg 253

Db 484 TAGCAACACCCAGAGTGTCTGCCTGTTGATGTCCGGAGCAGGAAGTTCAGGAAAG 430

RESULT 10

CB773455

LOCUS CB773455

DEFINITION AMGNNUC:TRPG2-00003-G7-A trpg2 (10294) Rattus norvegicus cDNA clone

trpg2-00003-g7 5', mRNA sequence.

ACCESSION CB773455

VERSION CB773455.1 GI:29861846

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 396)

AUTHORS Amgen EST Program.

TITLE Amgen Rat EST Program

JOURNAL Unpublished (2003)

COMMENT Contact: Dan Fitzpatrick

Amgen, Inc

One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA

Tel: 805 447-4881

Plate: 0003 row: g column: 7.

FEATURES

source

1..396

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="trpg2-00003-g7"

/tissue_type="pancreal gland brain"

/clone_lib="trpg2 (10294)"

/note="Vector: C6KFGF7L; Site_1: SalI; Site_2: NotI; pancreal gland brain region"

ORIGIN

Alignment Scores:

Pred. No.: 1.15e-40 Length: 396

Score: 564.00 Matches: 118

Percent Similarity: 69.77% Conservative: 2

Best Local Similarity: 68.60% Mismatches: 8

Query Match: 42.53% Indels: 44

DB: 14 Gaps: 2

US-10-078-090-151 (1-260) x CB773455 (1-396)

QY 55 AlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeu 74

Db 11 GCCCGGGTGACCCACGGGAGGCTGTCAAGGCTCTCAAGGGCTCCAAGAGTTGGTGCTG 70

QY 75 SerValTySerAlaGlyArgIleProGlyGlyTyThrValThrAsnHisIleTyThrTrp 94

Db 71 TCTGTATACTCAGTGGCGGTATACCGGGGGCTATGTCAACCAACCATCTACACCTGG 130

QY 95 ValAspProGlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGly 114

Db 131 GTGGACCCCGAGGGTCAAGACATCCCTCCCTCCAGCCTT-----CCCCAATGGCAGC 184

QY 115 AlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAsp 134

Db 185 ACCCTGACACAGCATGAAGATGATCGAAGGAGTGCCTTACACCTCCTGCAGAGTGGAGAT 244

QY 135 GluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGly 154

Db 245 GAGAAA----- 250

QY 155 GlyAlaGluTyrglyLeuGlyIleTyrlleThrGlyValAspProGlySerGluAlaGlu 174

Db 250 ----- 250

QY 175 GlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsn 194

Db 251 -----AAGTTGGGGACCAGATTCTGGAGGTGAACGGGGGAGCTTTCTCAGC 298

QY 195 IleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrVal 214

Db 299 ATCTGTCATGACGAGGCGAGTGAAGCTGCTCAAGTCATCCCGGCACCTCATCTGACCGTG 358

QY 215 LysAspValGlyArgLeuProHisAlaArgThrThr 226

Db 359 AAGGACGTGGGAGGCTGCCCCACGCACGTACCACC 394

RESULT 11

CD760225

LOCUS CD760225

DEFINITION CD760225 870 bp mRNA linear EST 30-JUN-2003

IMAGE:6961736 5', mRNA sequence.

ACCESSION CD760225

VERSION CD760225.1 GI:323444518

KEYWORDS EST.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

REFERENCE 1 (bases 1 to 870)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Leonard I. Zon, M.D.

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLAM14597 row: f column: 07

High quality sequence stop: 567.

FEATURES

location/Qualifiers

1..870

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="IMAGE:6961736"

/lab_host="DH10B (TI-resistant)"

/clone_lib="NCI CGAP_ZKId1"

/note="Organ: Kidney; Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Constructed by J. Wang (Research Genetics, Invitrogen Corp) from tissue donated by L. Zon (Harvard University). Note: this is a NCI CGAP Library."

ORIGIN

Alignment Scores:

Pred. No.: 3.1e-39 Length: 870
Score: 553.50 Matches: 118
Percent Similarity: 74.59% Conservative: 17
Best Local Similarity: 65.19% Mismatches: 45
Query Match: 41.74% Indels: 2
DB: 14 Gaps: 1

US-10-078-090-151 (1-260) x CD760225 (1-870)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 233 CTCAGCGTCACAAGAGCAATGAAGTTTGGGATTAGCATCCGCGGAGGATCAGAGCAT 292
QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 293 GGAGTTGGAATCTATGTTTCACTGGTGGAACTTGGATCACTGGCTGAAAGGAGGTTTG 352
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 353 AGAATGGGAGACCAATAATGAAAGTTAACGACAAAGTGTTCACCGGTCACGCACGCG 412
QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 413 GATGAGTAAAGGTGCTGAAAGCGACGAAAGCTCTGCATGTCGGTTCCTGTGGGC 472
QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 473 AGGATACAGCGGTTATATACCAATCACGTCTACACCTGGTGGATCCTCAGGTCGC 532
QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAla---LeuArgGlnGln 119
Db 533 AGTGTGTCTCTCCACCCGACCTGCTGGCGGAGCATCGCAGTCCACCTGCGCAGGTCC 592
QY 120 GluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsn 139
Db 593 AACAGTCAGGCGCGCAGTCACATGCAGCTACTGCAGGATGGAGATGAGAAGAACGTAAAT 652
QY 140 LeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGly 159
Db 653 CTGGTGTCTGGAGACTGCCGCTCTTTGGGTTTGTATGATCTCTCGGATGACTGATTACTT- 711
QY 160 LeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLys 179
Db 712 CTGGGTATCTTATTAAAGGTTGAAATAGTTTGCAGCGGATTTGAGGATAAGGGATAACT 771
QY 180 Val 180
Db 772 ATC 774

RESULT 12 CB692138 374 bp mRNA linear EST 10-APR-2003
LOCUS AMGNNUC:TRPG2-00004-D6-A trpg2 (10294) Rattus norvegicus cDNA clone
DEFINITION trpg2-00004-d6 5', mRNA sequence.

ACCESSION CB692138
VERSION CB692138.1 GI:29749285
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 374)
AUTHORS Amgen EST Program.
TITLE Amgen Rat EST Program
JOURNAL Unpublished (2003)
COMMENT Contact: Dan Fitzpatrick
Amgen, Inc

One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00004 row: d column: 6.
Location/Qualifiers
1. 374

FEATURES source

/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="trpg2-00004-d6"
/tissue_type="peneal gland brain"
/clone_lib="trpg2 (10294)"
/note="Vector: C6KFGF7L; Site_1: SalI; Site_2: NotI;
peneal gland brain region"

ORIGIN

Alignment Scores: 2.67e-37 Length: 374
Pred. No.: 526.00 Matches: 111
Score: 526.00
Percent Similarity: 68.48% Conservative: 2
Best Local Similarity: 67.27% Mismatches: 8
Query Match: 39.67% Indels: 44
DB: 14 Gaps: 2

US-10-078-090-151 (1-260) x CB692138 (1-374)

QY 55 AlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeu 74
Db 11 GCCCGGTGACCCACGCGGAGGCTGTCAAGGCTCTCAAGGCTCCAAGAGTTGGTGTG 70
QY 75 SerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp 94
Db 71 TCTGTATACTCAGCTGGGCGGTATACCCGGGGGCTATGTCAACCAACCATCTACACCTG 130
QY 95 ValAspProGlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGly 114
Db 131 GTGGACCCCGAGGTCGAAGCACATCCCTCCCTCCAGCCTT-----CCCCATGGCAGC 184
QY 115 AlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAsp 134
Db 185 ACCCTGAGACACATGAAGATGATCGAAGGAGTGCCTACACCTCTCGCAGAGTGGAGAT 244
QY 135 GluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGly 154
Db 245 GAGAAA-----
QY 155 GlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGlu 174
Db 250 -----
QY 175 GlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsn 194
Db 251 -----AAGGTTGGGACCAGATTCTGGAGGTGAACGGCGGAGGCTTTCTCAGC 298
QY 195 IleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuLeuThrVal 214
Db 299 ATCTGTCATGACGAGGAGTGAAGCTGCTCAAGTCATCCCGGCACCTCATCTCTGACCGGTG 358
QY 215 LysAspValGlyArg 219
Db 359 AAGGACGTCGGGAGG 373

RESULT 13 CB693399

LOCUS CB693399

DEFINITION AMGNNUC:SRPG2-00029-A8-A srpg2 (10238) Rattus norvegicus cDNA clone

srpg2-00029-a8 5', mRNA sequence.

ACCESSION CB693399

VERSION CB693399.1 GI:29750546

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 343)

AUTHORS Amgen EST Program.

TITLE Amgen Rat EST Program

JOURNAL Unpublished (2003)

CB693399 343 bp mRNA linear EST 10-APR-2003
AMGNNUC:SRPG2-00029-A8-A srpg2 (10238) Rattus norvegicus cDNA clone
srpg2-00029-a8 5', mRNA sequence.

CB693399
CB693399.1 GI:29750546
EST.
Rattus norvegicus (Norway rat)

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 343)

Amgen EST Program.

Amgen Rat EST Program

Unpublished (2003)

COMMENT Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00029 row: a column: 8.
Location/Qualifiers
FEATURES
source
1. .343
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="srp92-00029-a8"
/tissue type="peneal gland brain"
/clone_lib="srp92 (10238)"
/note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; peneal gland brain region"

ORIGIN
Alignment Scores:
Pred. No.: 4.37e-37 Length: 343
Score: 523.00 Matches: 102
Percent Similarity: 90.27% Conservatve: 0
Best Local Similarity: 90.27% Mismatches: 9
Query Match: 39.44% Indels: 2
DB: 14 Gaps: 1

US-10-078-090-151 (1-260) x CB693399 (1-343)

QY 81 ArgileProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100
Db 10 CGTATACCCGGGGCTATGTACCAACACCATACACCTGGGTGGACCCCGAGGTCGA 69

QY 101 SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 70 AGCATATCCCTCCCTCCAGCCTT-----CCCATGGCAGCACCCCTGAGACGATGAA 123

QY 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGlyLysLysValAsnLeu 140
Db 124 GATGATCGAAGGAGTGCCCTACACCTCCTGCAGAGTGAGATGAGAAAGGTGAACCTG 183

QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
Db 184 GTGTGGGGGACGGCCGCTCTGGCCCTACAATCCGAGGTGGAGCCGAGTACGGCCTT 243

QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
Db 244 GGCATTATACATCACTGTGTGGACCCCGGCTCTGAAGCCGAAAGCAGCGGCCTCAAGTT 303

QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
Db 304 GGGGACCAGATTCTGGAGGTGAACGGCGGAGCTTTCTC 342

RESULT 14
BB872069
LOCUS BB872069 383 bp mRNA linear EST 27-NOV-2001
DEFINITION BB872069 RIKEN full-length enriched, 1 month neonate cerebellum Mus musculus cDNA clone G630034G23 5', mRNA sequence.
ACCESSION BB872069
VERSION BB872069.1 GI:171118279
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 383)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
Location/Qualifiers
1. .383
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G630034G23"
/tissue type="cerebellum"
/dev_stage="1 month neonate"
/clone_lib="RIKEN full-length enriched, 1 month neonate cerebellum"

ORIGIN
Alignment Scores:
Pred. No.: 1.17e-36 Length: 383
Score: 519.00 Matches: 103
Percent Similarity: 97.17% Conservatve: 0
Best Local Similarity: 97.17% Mismatches: 3
Query Match: 39.14% Indels: 0
DB: 10 Gaps: 0

US-10-078-090-151 (1-260) x BB872069 (1-383)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlySerGluHis 20
Db 66 CTGCGGCGCGCCCAAGCCCAAGCGGCTTGGGCTTCAGCATCCGGGGGGGCTCGAACAG 125

QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGlyLysGlyLeu 40
Db 126 GGGGTGGGCATCTACGTGTCTCTAGTGGAGCCGGGCTCCCTGGCAGAGAAGGGTTG 185

QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 186 CGGGTCCGGGACCAAGATTCTGCGCGTCAACGATAAATCTCTAGCCCGGTGACCCACGCG 245

QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 246 GAGGCTGTCAAGGCTCTCAAGAGGCTCCAAAGGCTCCAGAGGCTGTCTGTGTACTCAGCTGGG 305

QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100

Db306CGTATCCAGGGGGCTATGTGACCAACCAAAATCTACACCTGGGTGGACCCACAGGTGCA365

Qy101SerIleSerProProSer106
|||||

Db366AGCACATCCCCTCCCTCC383

RESULT15

BF953476/c

LOCUS

DEFINITIONRC3-NN1182-121100-011-g08 NN1182 Homo sapiens cDNA, mRNA sequence.

ACCESSIONBF953476

VERSIONBF953476.1GI:12370751

KEYWORDS

SOURCEEST.

ORGANISMHomo sapiens (human)

ORGANISMHomo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 325)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNALProc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE20202663

PUBMED10737800

COMMENTContact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC3-NN1182-121100-011-g08&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 325.

FEATURES

source

1..325

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="NN1182"

/note="Organ: nervous normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

3.9e-36

512.00

99.03%

99.03%

38.61%

10

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

325

102

0

1

0

0

US-10-078-090-151 (1-260) x BF953476 (1-325)

Qy125SerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeuValLeuGlyAsp144
|||||

Db320AGCACCTGCACCTCCTGCAAGAGGGGATGAGAAAAAGGTGAACCTGGTCTGGGGGAC261